ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

Evaluating and improving the risk adjustment scheme in Switzerland

Identifying predictable spending variation within morbidity groups

Author:	Jules Smeets (454567)
Supervisor:	dr. Richard van Kleef
Second assessor:	prof. dr. Owen O’Donnell
Date:	6 July 2023
Wordcount:	13185

Abstract

This paper analyses to what extent selection incentives exist within Pharmaceutical Cost Groups in the Swiss risk adjustment scheme and to what extent these selection incentives can be reduced by supplementing the Swiss risk adjustment scheme with outlier risk sharing. I use data from a Swiss health insurer (N = 805,531) to replicate the Swiss risk adjustment scheme. Using a Random Forest model, I estimate predictable profits and losses for individuals in morbidity groups. I conclude that considerable selection incentives exist within Pharmaceutical Cost Groups. In addition, I find that outlier risk sharing can reduce predictable losses and profits within morbidity groups. These reductions are largest for individuals with the highest predictable profits and losses.

Table of contents

Acknowledgements

I would like to gratefully thank dr. Richard van Kleef for his valuable supervision during this thesis project. I could not have wished for a better supervisor. I would also like to thank Celine for her unwavering support, Mats for this editorial support and feedback, and Joar for the many hours we spent together during the thesis process. I would also like to express my gratefulness to CSS Health Insurance for the permission to use their administrative data for the purpose of this thesis.

Chapter 1: Introduction

In 1991, The federal council of Switzerland proposed the *"Loi fédérale sur l'assurance-maladie".* This officially put Switzerland on the road toward a healthcare system based on regulated competition. Today, Switzerland has compulsory health insurance for all adults, a community-rated premium and periodic open enrolment. All enrolees can change their insurer once a year, leading to competition among insurers (Schmid, Beck, & Kauer, 2018). To provide the best price and quality to the insured, insurers can increase efficiency and act as prudent buyers by using managed care techniques such as selective contracting of providers and by using alternative payment models. This can, at least in theory, reduce costs and increase the quality of healthcare. Switzerland's system is often seen as exemplary for regulated competition in healthcare.

However, a major challenge faced by countries with regulated competition is risk selection. Community-rated premiums create predictable profits and losses for certain individuals. For example, a young and healthy individual has lower expected healthcare expenditures than a 70-year-old diabetes patient. This incentivizes insurers to enrol as many healthy people as possible while trying to keep out chronically ill and otherwise unhealthy individuals. For example, if people with heart disease are predictably unprofitable, a health insurer might be incentivized to contract only low-quality healthcare providers for heart disease. This way, the health insurer is unattractive to patients with heart disease. Consequently, they are less likely to contract with this health insurer. Thus, the health insurer can increase its profits by purposefully contracting with the worst healthcare providers. This is often seen as undesirable.

To mitigate risk selection, Switzerland introduced a risk adjustment scheme parallel to the community-rate premiums. As of this moment, the variables in the Swiss risk adjustment model are gender, age, an indicator of hospitalization in the previous year, and several pharmaceutical cost groups (PCG). If patients receive a certain amount of drugs in the previous year that are associated to a particular disease, they are assigned to a PCG that is associated with that disease. The PCGs are then used as predictors of healthcare expenditure (Beck, Kauer, McGuire, & Schmid, 2020). For example, if an individual used more than the threshold amount of insulin in the previous year, he or she will be assigned to the diabetes PCG in the next year.

Unfortunately, research has shown that even under the most sophisticated risk adjustment models, predictable profits and losses still exist. This means there are still incentives for risk selection. This gives societal relevance to the research done in this field. In this paper, I build on previous research by zooming in on PCGs. While they account for the difference in spending *between* groups of people, the key novelty in this paper lies in the fact that I analyse to what extent variation in healthcare spending is predictable *within* PCGs. More specifically, I will examine to what extent residual spending within PCGs is predictable. By residual spending, I mean spending net of risk adjustment i.e. actual spending minus predicted spending according to the risk adjustment model.

Any predictable variation in residual spending found within PCGs can lead to selection incentives. To limit these incentives, one solution could lie in outlier risk sharing. In contrast to risk adjustment, which is based on *predicted* healthcare costs that are estimated *beforehand*, risk sharing is based on *actual* healthcare costs that are calculated *afterwards*. One form of risk sharing is when an insurer is reimbursed for very high-cost individuals above a certain threshold. An example would be when for every individual, 100% of costs above CHF 50.000 is reimbursed. This is called outlier risk sharing (McGuire & van Kleef, 2018). After identifying the predictable variation in residual spending within PCGs, I will try to reduce this by supplementing the risk adjustment model with outlier risk sharing.

Chapter 2: Objective & Research question

This paper aims to analyse predictable variation in residual spending within PCGs and aims to explore whether any predictable variation in residual spending can be reduced by implementing outlier risk sharing. This results in the following research question:

"To what extent do selection incentives exist within Pharmaceutical Cost Groups of the Swiss Risk Adjustment Scheme and to what extent can outlier risk sharing help reduce such selection incentives?"

Several sub-questions have been formulated to act as steppingstones to answer the main research question:

1) *What are the characteristics of people in a PCG in terms of mean spending, age, gender, and health compared to those who are not in a PCG?*
2) *To what extent is variation in residual spending within the group of people with a PCG predictable?*
3) *To what extent can outlier risk sharing help reduce predictable variation in residual spending within the group of people with a PCG?*

Even the most sophisticated risk adjustment models cannot eradicate all predictable profits and losses (van Kleef, Eijkenaar, & van Vliet, Selection Incentives for Health Insurers in Sophiticated Risk Adjustment, 2020). As a result, research in this field remains desirable. After all, a better risk adjustment model leads to fewer risk selection incentives. The scientific novelty in this paper is the focus on the existence of predictable profits and losses among the chronically ill after the application of risk adjustment. If predictable profits and losses continue to exist among the chronically ill after risk adjustment, health insurers might have incentives for risk selection. After successful risk adjustment, health insurers should have an incentive to provide the best care to all patients.

In the next chapter, a theoretical background will be provided for the main concepts that will be used in this paper. In chapter four, I will discuss the methodological strategy that was used in this paper. Chapter five contains an overview of the results. In chapter 6, a summary of the key findings will be provided. and the results will be put in a broader perspective.

Chapter 3: Theoretical background

In this chapter I will provide some theoretical background for the main concepts that will be used in this paper. First, I will discuss regulated competition and the role of risk adjustment. Second, I will focus on the Swiss risk adjustment model with particular attention for PCGs. Third, methods for predicting healthcare costs will be discussed, with special attention to the Random Forest model. Fourth, an overview of performance indicators will be provided for quantifying model performance and selection incentives. Lastly, risk sharing will be discussed.

3.1. Regulated competition and the role of risk adjustment

The idea of regulated competition can be traced back to Alain Enthoven. While his ideas have evolved and changed over time, the key feature of Enthoven's model is the existence of a "sponsor" on the demand side of the healthcare market. This sponsor actively manages the health plan market to overcome market failure (Enthoven, 1978). In European countries, this sponsor is mostly the government. However, in the USA, some employers also function as such.

In 1993, Enthoven defined managed competition as a purchasing strategy to maximize value of money for consumers, based on rational microeconomic principles. Where a Sponsor must design the rules of competition so *"as not to reward health plans for selecting good risks, segmenting markets, or otherwise defeating the goals of managed competition"* (Enthoven, 1993).

Most systems of regulated competition have some combination of mandatory health insurance, community-rated premiums, open enrolment, and standard benefits packages. The goal is to increase the availability, affordability and fairness of care (van de Ven, et al., 2013). This creates a system based on solidarity, with cross-subsidies from the low risks to the high risks. However, systems based on community-rated premiums also result in predictable profits for the healthy, and predictable losses for the unhealthy. This creates incentives for risk selection (Arrow, 1963). For example, insurers could focus their marketing and customer service on young people. In addition, they could decide not to contract with providers who are relatively attractive to unprofitable consumers. By making their products undesirable for unprofitable consumers, an insurer might be able to provide a lower premium than its competitors. This way, differences in premiums would not only reflect differences in healthcare quality and efficiency but also differences between the initial health of the participants of health insurance (van Kleef, Eijkenaar, & van Vliet, Selection Incentives for Health Insurers in Sophiticated Risk Adjustment, 2020).This is often seen as undesirable by policymakers.

To reduce incentives for risk selection, risk adjustment schemes were first proposed in the 1980s (Newhouse, 1986). Risk adjustment reduces predictable profits and losses by compensating healthcare payers based on predicted healthcare spending. Risk adjustment can be summarized as *"a bundled payment strategy in which payments are based on mathematical formulas that predict health plan obligations for spending on each enrolee"* (Ellis, Martins, & Rose, 2017). Research has shown that these schemes can significantly reduce predictable profits and losses, but still fail to eliminate them (van Kleef, Eijkenaar, & van Vliet, Selection Incentives for Health Insurers in Sophiticated Risk Adjustment, 2020).

3.2. The Swiss risk adjustment model design and the role of PCGs

As of 2023, the Swiss model is a zero-sum system with no transfers between cantons (provinces). This means per canton, insurers with a below-average risk profile must pay into the risk adjustment fund, and insurers with an above-average risk profile receive money from the fund. Risk adjustment takes place for all individuals except for people younger than 18 years and asylum seekers. The risk adjusters are gender, 15 age groups, an indicator for at least three consecutive nights in hospital in the previous year, and several PCGs (Schmid, Beck, & Kauer, 2018).

More specifically, gender and age groups are combined in a set of 30 interaction terms to allow for differences in the lifetime trends of expected costs for males and females. (Kauer, McGuire, & Beck, 2020) One example of a difference in lifetime trend is the expected health care costs of pregnancy during the fertile ages, which are higher for women. The indicator for hospitalization is a dummy variable. It takes the value of one if an individual was hospitalized for at least three consecutive days in the previous year and the value zero otherwise. The PCGs are also dummy variables. They take the value one of an individual used a certain threshold of medication in the previous year and the value zero otherwise.

The introduction of the PCGs in 2020 was the last major improvement of the Swiss risk adjustment model. As discussed briefly, PCGs are based on drug use. Drugs are categorized based on their active ingredient (and can therefore be associated with specific illnesses). Inclusion in a PCG is dependent on passing a threshold of drug use based on standardized daily use (Kauer, McGuire, & Beck, 2020). For example, if an individual uses a large amount of insulin in the previous year (higher than the threshold value), he or she will be assigned to the PCG for diabetes patients. In total, the Swiss risk adjustment model has 28 PCGs. Before the introduction of PCGs, the estimated R-squared of the Swiss risk adjustment model sat between 21.0% and 12.2%. Since the introduction of PCGs, the R-squared estimations range between 30.0% and 21.2%. This shows a remarkable improvement in predictive power (Schmid, Beck, & Kauer, 2018).

More broadly speaking, the predictive power of risk adjustment models has been increasing in many countries through the introduction of health-based risk adjusters like PCGs, diagnostic cost groups (DCGs), and indicators for historical healthcare expenditure (van de Ven, Beck, van de Voorde, Wasem, & Zmora, 2007). This has compensated large parts of the predictable variance in spending between different groups of people. However, not much is known about the variance within the subgroups created in risk adjustment models or the potential risk selection incentives that exists within such groups. This paper intends to shed light on this issue by examining to what extent the variation in spending net of risk adjustment is predictable within the group of people with at least one PCG.

3.3. Methods for predicting healthcare costs

3.3.1. Regression models and machine learning

For predictive modelling of healthcare expenditure, many different models such as OLS, GLM, two-part models and machine learning techniques have been used (Ellis, Martins, & Rose, 2017). OLS is widely used in risk adjustment. However, Generalized Linear Models (GLM) can be more flexible in the assumed distribution of the dependent variable. This is interesting considering healthcare expenditure is often highly skewed. One type of GLM that could be useful is a two-part model. This is well-suited for modelling data containing many zeros by combining a first-stage binary regression with a second-stage regression of the non-zero values (Powers, Meyer, Roebuck, & Vaziri, 2005).

In addition, the advent of machine learning has brought potentially more predictive models to the field. An advantage of machine learning is that it can find interactions that would otherwise have remained unknown without pre-existing knowledge (Ellis, Martins, & Rose, 2017). A growing amount of evidence suggests that machine learning is well suited to predicting healthcare expenditure. One study trained several regression trees on residual spending variation within the Dutch Risk adjustment system. The newly discovered interaction terms were included in the original Dutch risk adjustment model, resulting in improved measures-of-fit. (van Veen, van Kleef, van de Ven, & van Vliet, 2018). A study in France showed that machine learning models like Neural Network (NN) and Random Forest

(RF) outperformed GLM consistently (Vimont, Leleu, & Durand-Zaleski, 2022). Another study on risk adjustment for oncology expenditure found that machine learning yielded better results than GLM, with Random Forest showing the best performance (Mazumdar, 2020).

The main disadvantage of machine learning models is that they do not provide a transparent set of weighted predictors and are therefore difficult to interpret. Random forest, for example, is based on the aggregate value of many regression trees and could be described as a 'black box' algorithm (Biau & Scornet, 2016). For risk adjustment models, fairness and transparency are usually valued next to predictive value. This makes it difficult to replace GLMs used for actual risk adjustment. In addition, when the aim is to find new or better predictors to add to an existing risk adjustment model, there is also a need to be able to interpret relationships in the data. This narrows the scope in which "black box" algorithms can be used.

However, machine learning can still be relevant when an explicit model with defined relationships is not necessary. In this case, the main objective is to accurately predict variation in residual spending in PCGs. Interpretation of the relationships is therefore less relevant. Thus, when a black box algorithm can provide robust estimates with higher measures-of-fit than GLMs, this could be very useful.

3.3.2. Random Forest and Regression Trees

The Random Forest Algorithm was first introduced by L. Breiman in 2001 and is based on so-called "classification and regression trees" (CART). For simplicity, I will call them regression trees. (Breiman, 2001). A regression tree "grows" by splitting a sample into more and more subsamples. It does this by finding the best possible split for every (sub)sample. A split is better when it results in more homogeneous subsamples. This way a metaphorical tree grows in which all the subsamples are "nodes" from which two or more branches reach out to the next node. All subsamples are split until they are fully homogeneous (or when told to stop earlier). These final subsamples are called "leaves". The higher a tree grows, the higher order interaction terms it uses (van Veen, van Kleef, van de Ven, & van Vliet, 2018). To make a prediction, a tree will compare the values of new data with the splitting criteria at every node. Data will traverse down the tree's branches until a final leaf is reached.

Regression trees have been found to be very useful in finding unknown relationships and interaction effects. However, they are prone to overfitting. This is when a regression tree describes random variation in the sample rather than true relationships (Hastie, Friedman, & Tibshirani, 2009). The aim of a Random Forest is to maintain the usefulness of regression trees while reducing the issues of overfitting. The issue of overfitting will be discussed in more detail in paragraph 3.4.2.

A Random Forest model is based on a (large) number of regression trees. First, a bootstrapped sample is created for every tree in the Random Forest. Bootstrapping is done by recreating a sample using randomly drawn observations from the original sample. Second, a limited number of variables are randomly selected for every bootstrapped sample. The regression tree is only allowed to base its splits on the selected variables. Third, a regression tree is grown for all the bootstrapped samples based only on the randomly selected variables. Predictions of the Random Forest are based on the aggregate output of every tree. This sequence of bootstrapping and aggregating is called "bagging". This makes the estimate more robust to overfitting. The rationale is that for every tree that describes random variation, there will be another tree that describes random variation in the opposite direction (Biau & Scornet, 2016).

3.4. Measures for quantifying model performance and selection incentives

3.4.1. Measures for quantifying model performance

To measure the performance of a Random Forest model or any other predictive model for risk adjustment, several "measures-of-fit" are generally used. Broadly speaking, there are three types of measures-of-fit. Namely, measures based on (1) squared errors; (2) absolute errors, and (3) untransformed errors (van Veen, van Kleef, van de Ven, & van Vliet, 2015). All types have their advantages and disadvantages, making it useful to judge the predictive value of a risk adjustment model based on several different types of measures-of-fit. The first two will be discussed in this paragraph, the third will be discussed in the next paragraph.

Measures based on squared errors are most commonly used. Examples are *R-squared* and the *Mean Squared Prediction Error* (MSPE). These are based on the squared value of the prediction errors. This ensures that both over and underestimations are considered and that greater estimation errors receive a greater value. R-squared can be calculated as follows (van Kleef, McGuire, van Vliet, & van de Ven, 2017):

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2} \qquad (1)$$

In this formula, Y_i denotes the actual residual spending for individual i, $\hat{Y}_i$ denotes the predicted residual spending for individual i, and $\bar{Y}$denotes the average residual spending. R-squared has a value between 0 and 1 where a higher score indicated better performance (Ash, McCall, Fonda, Hanchate, & Speckman, 2005).

The *Cummings' Prediction Measure* (CPM) is an example of a measure based on absolute prediction errors. Using absolute values means that positive and negative errors do not cancel each other out. In addition, as the errors are not squared, the measure is less sensitive to very high or very low values. It is calculated as follows (van Kleef, McGuire, van Vliet, & van de Ven, 2017):

$$CPM = 1 - \frac{\sum_{i=1}^{n}\left|Y_i - \hat{Y}_i\right|}{\sum_{i=1}^{n}\left|Y_i - \bar{Y}\right|} \qquad (2)$$

The nominator represents the sum of the absolute values of the difference between actual and predicted spending for all individuals in the sample. The denominator represents the sum of absolute values of the difference between the actual spending for every individual and the average actual spending in the sample. (Ash, McCall, Fonda, Hanchate, & Speckman, 2005). This results in a value between 0 and 1, where a higher value represents better performance, similar to R-squared.

3.4.2. Measures for quantifying selection incentives

While measures like R-squared and CPM can be useful indicators of a models' performance, they cannot immediately show the presence of selection incentives. For example, an R-squared score does not tell us what groups are profitable and what groups are unprofitable. To gain insights into the presence of selection incentives, measures based on untransformed errors can be used. One example of an intuitive measure based on untransformed errors is over- or under compensation. This is calculated using the following formula (Layton & Ellis, 2018):

$$Over/Undercompensation = \frac{\sum_{i \in g}\left(\hat{Y}_i - Y_i\right)}{n_g} \qquad (3)$$

When used in an OLS regression on a full sample, over and under compensation will be close to zero, meaning the sum of the predicted values is equal to the sum of actual values. It is therefore useful to use this measure on subsamples (van Veen, van Kleef, van de Ven, & van Vliet, 2015). Here the nominator represents the sum of residual spending for all individuals within the subsample of interest. The denominator is the number of individuals in that subsample. This essentially gives us the average residual spending for a specific subgroup.

3.4.2. Overfitting and the split-sample approach

When evaluating the performance of prediction models, one needs to consider the risk of overfitting. Overfitting has been shortly discussed in paragraph 3.3.1. It is caused by random variation in the data (noise) being interpreted as having predictive value (Ying, 2019). An indication of overfitting is when the performance of a model drops when confronted with unseen data. Let's say an individual i engages in curling. In fact, he is one of the only individuals in the dataset that does this. In addition, by complete coincidence, individual i also has very high healthcare expenditure. A machine-learning model might associate curling with very high healthcare costs. The model has mistaken random noise for an association between curling and healthcare expenditure and has overfitted. If this model is confronted with a new dataset, in which curling players do not have very high healthcare costs, the performance indicators will suddenly drop.

To improve the robustness of estimates to overfitting, a "split sample method" might be preferable. Here, the model is trained on one part of the sample and tested on the other. Measures-of-fit on the testing data will be more robust to overfitting. Large differences in measures-of-fit between the training and testing data indicate that the model was overfitted to the original data. (Faraway, 2016).

3.5. Risk sharing as a supplement to risk adjustment

As discussed, risk adjustment often fails to adequately compensate for all predictable profits and losses, especially for a small number of very high-cost individuals (van Kleef, Eijkenaar, & van Vliet, Selection Incentives for Health Insurers in Sophiticated Risk Adjustment, 2020). As a result, it might be attractive to look at alternative solutions. Especially when the possibilities for improving risk adjustment are limited, for example due to a lack of data. One alternative way to reduce predictable profits and losses is risk sharing. This is a payment system that redistributes funds among insurers based on actual cost (van Kleef, Reuser, Stam, & van de Ven, 2022).

In general, there are four types of risk sharing. The first and simplest form is *"proportional risk sharing"*. In this case, the regulator bears a fixed percentage of the healthcare costs. The second form is "*reinsurance*". Here, an insurer is itself insured against very high healthcare costs. The reinsurer could be the government, but it could also be a private entity. A reinsurer could for example agree to bear the healthcare costs of a patient for 80% above a threshold of CHF 300.000, - (as is the case in Switzerland), possibly in exchange for a fee. This type of risk sharing is also referred to as "*outlier risk sharing".* The third type of risk sharing is called *"high-risk pooling".* Here, the idea is to protect insurers against a group of high-risk enrolees by pooling them into a high-risk group (HRG) that is separated from the rest of the market. The fourth and last type is a *"risk corridor".* Here, a range is defined in which the insurers bear all risks. Outside of this range, the regulator will share a part of the risk. An example of a one-sided risk corridor is when the regulator bears all costs for a patient that exceed 110% of the revenues for the insurer (McGuire & van Kleef, 2018).

While risk sharing can decrease predictable profits and losses, it might also affect one of the fundamental principles of regulated competition: risk-bearing insurers. The fact that insurers bear the costs made for their enrolees incentivizes them to act cost-conscientiously and buy efficient care. If part of this risk is taken away by risk sharing, this might lead to fewer incentives for cost containment.

This leads to a risk sharing trade-off between risk selection incentives on the one hand, and cost containment incentives on the other hand (McGuire & van Kleef, 2018).

Currently, the Swiss health insurance scheme includes two types of cost sharing. First, proportional risk sharing in which roughly 55% of inpatient hospital care is not paid by the insurers, but directly by the cantons. Second, reinsurance. Healthcare insurers are obliged to buy reinsurance to preserve their solvency (Schmid, Beck, & Kauer, 2018).

In this chapter, a theoretical foundation was provided for the key concepts utilized in the paper. First, regulated competition and the significance of risk adjustment was examined. Next, the Swiss risk adjustment model was explored with special attention to PCGs. Additionally, various methods for predicting healthcare costs were discussed, with a specific emphasis on the Random Forest model. Next, an overview of performance indicators was provided to assess model performance and selection incentives. Finally, the topic of risk sharing in healthcare was discussed.

Chapter 4: Empirical strategy

4.1 Study Design

This study is designed to answer the research questions by quantitative methods. This is done by using microdata from a Swiss health insurer from 2014 to 2016. First, descriptives for PCG patients will be compared to non-PCG patients. Second, the Swiss risk adjustment model will be replicated to generate residual spending (actual healthcare spending minus predicted healthcare spending). Third, a model will be estimated to predict the extent to which variation in residual spending within PCGs is predictable. Lastly, an attempt will be made to lower predictable variation in residual spending within PCGs by supplementing the risk adjustment model with outlier risk sharing.

4.2 Data

The data used for this research has been made available by CSS. This is a Swiss health insurer and the current market leader with more than 1.6 million enrolees. The dataset consists of microdata for more than 1 million enrolees from 2014 to 2016. The most important variables are listed in table 1:

Table 1: Description of key variables

Variables	***Description***
Age groups	Age group of individual i in year t. There are 14 age groups in total: 0-18, 19-25, 26-30, 31-35, …., 80+.
Gender	Gender of individual i.
Total healthcare spending	Total gross healthcare spending for individual i in year t.
Hospitalization in the prior year	Indicator whether individual i was hospitalized for 3 or more consecutive nights in year t – 1.
PCG groups	A set of 28 dummy variables indicating whether individual i is classified in one of the PCGs in year t. PCGs are based on the use of prescription drugs in year t-1.

4.3 Methodology

Below, I discuss the methodological steps that were taken in this paper. First, descriptive statistics were compared between individuals who were assigned to at least one PCG (the PCG group) and individuals who were not assigned to a PCG (no PCG group). Second, the Swiss risk adjustment model was replicated to obtain residual spending variation within the PCG group. Third, the predictability of residual spending was estimated using a Random Forest model. Fourth, the Swiss risk adjustment model was supplemented with outlier risk sharing and step three was repeated. Finally, the predictability of residual spending within the PCG group was compared between the risk adjustment model with and without outlier risk sharing.

4.3.1. Data transformations

Before the results could be estimated, some data transformation was necessary. First, the cross-sectional datasets for all years were combined into a single panel dataset. Second, individuals who are not present in all years were removed from the sample. This was needed to maximize the predictive power of the Random Forest model, as it does not allow for missing values. Third, a dummy variable was created that takes the value of one if an individual is assigned to any of the PCGs and zero otherwise. This was done because using separate PCGs as the basis for the analyses yielded a small number of observations, decreasing statistical power. Using all individuals who are assigned to at least one PCG as one group maximizes the number of observations. Lastly, a set of dummy variables was created containing interaction terms of age groups and gender. This was necessary to replicate the Swiss risk adjustment model.

4.3.2. Descriptive statistics

First, descriptives are provided for the full sample. Second, descriptives are provided and compared between PCG-assigned individuals and non-PCG-assigned individuals. This provides insight into the association between being assigned to a PCG and other socio-economic variables.

4.3.3. Replicating the Swiss risk adjustment model and obtaining residual spending

While the real-life risk adjustment model is separated for every canton, I used a single model for the whole of Switzerland to maximize the number of observations used per regression. This means I must assume that the variation in residual spending does not differ between cantons. This will be further discussed in chapter 6. The Swiss risk adjustment model was recreated using an OLS with the total healthcare expenditure of individual i in 2016 as the dependent variable. As independent variables, several dummy variables were used. First, a set of interaction terms between gender and age groups; second, a dummy variable indicating hospitalization in the previous year and third, dummy variables for every PCG. Using the estimated healthcare expenditure per individual, I then obtained residual spending by subtracting the actual healthcare spending from the estimated healthcare spending. Residual spending can be interpreted as losses per individual after risk adjustment. This gives us the following formulas:

Estimated healthcare spending: $$\hat{Y}_i = b_0 + b_1X_{1i} + b_2X_{2i} + b_3X_{3i} + .. + b_{56}X_{56i} \quad (5)$$

Residual spending: $$R_i = Y_i - \hat{Y}_i \quad (6)$$

Table 2: Variable definitions

$\hat{Y}$	The estimated healthcare expenditure for individual i in 2016.
Y	The actual healthcare expenditure for individual i in 2016.
b_0	The intercept. This can be interpreted as the mean healthcare expenditure when all the predictor variables are equal to zero
$b_1X_1 + b_2X_2 + b_3X_3 + .. + b_{56}X_{56}$	These are all the predictor variables. They consist of: (1) A set of 28 interaction terms between gender and age group at time t; (2) An indicator if individual i was hospitalized for at least three consecutive days in year t-1 and (3) A set of dummy variables for all PCGs, based on pharmaceutical use in year t-1. This includes a dummy for individuals are not assigned to any PCG.

4.3.4. Determining the predictability of residual spending variation within PCGs

The next step is to examine the predictability of the residual spending within PCG groups. Based on a review of the literature, a Random Forest model was used to do this. There are several reasons for using Random Forest. First, machine learning models can find previously unknown and potentially complicated interaction effects, unlike traditional GLMs. Second, the literature shows that machine learning models consistently outperform traditional GLM models in predicting healthcare expenditure. Random Forest seems to perform particularly well in this field. For more on this, I refer to paragraph 3.3.

Random Forest

The dependent variable in the Random Forest model is the residual healthcare spending for individual i in 2016. In contrast to the Swiss risk adjustment model, which is based solely on data from 2016, the Random Forest model can also use data from previous years. I am only interested in predictable healthcare spending, so only data was used that was known before the start of 2016. Therefore, the following independent variables were used: (1) The full set of age groups for 2014, 2015 and 2016 (2) gender, (3) an indicator for hospitalization for 2014, 2015 and 2016 (4) the full set of dummy variables for PCGs for 2014, 2015 and 2016 and (5) healthcare spending in 2014 and 2015.

Several parameters had to be determined to tune the Random Forest model, namely the number of regression trees and the number of variables considered per node (mrty). The following steps were taken to achieve the optimal value for these parameters. First, a default model was constructed with 100 trees and 9 mrty. The number of trees was deliberately chosen to be relatively high, so the optimal number of trees could be determined based on the results of the default model. The value of 9 mrty was based on the rule of thumb that mrty should be roughly equal to the square root of the number of x variables (√89 = 9.43) (Biau & Scornet, 2016). For training the default model and tuning models, a subset of 10,000 randomly chosen observations was used and split in 8000 training observations and 2000 testing observations. The main disadvantage of tuning the RF model on a subsample instead of the full sample is that not all information is used. This means that there is a chance that not all chosen parameters are optimal when used on the full sample. However, it provides computational efficiency. The final model was trained on the full sample, which was divided in to 30% training data and 70% testing data.[1]

[1] In the literature the data is mostly split to a ratio of 80% training data and 20% test data (Nguyen, et al., 2021). However, after extensive testing computational capacity seemed to be reached with a maximum training dataset of 30% and a testing dataset of 70%.

Tuning mtry

The value of mrty was tuned in two steps. First, a random search method was used. A random search is a comparison of the predictive scores of several models with randomly chosen parameters within a certain range. For this random search, ten random mtry values between 3 and 60 were chosen. Next, ten different RF models were run with the randomly chosen mtry values. The results can be seen in table 3. The highest predictive score was achieved by 43 mrty, and predictive power seems to be increasing with the number of variables considered. However, it is important to note that higher levels of mtry will increase the risk of overfitting. In addition, 13 mrty seemed to score relatively high.

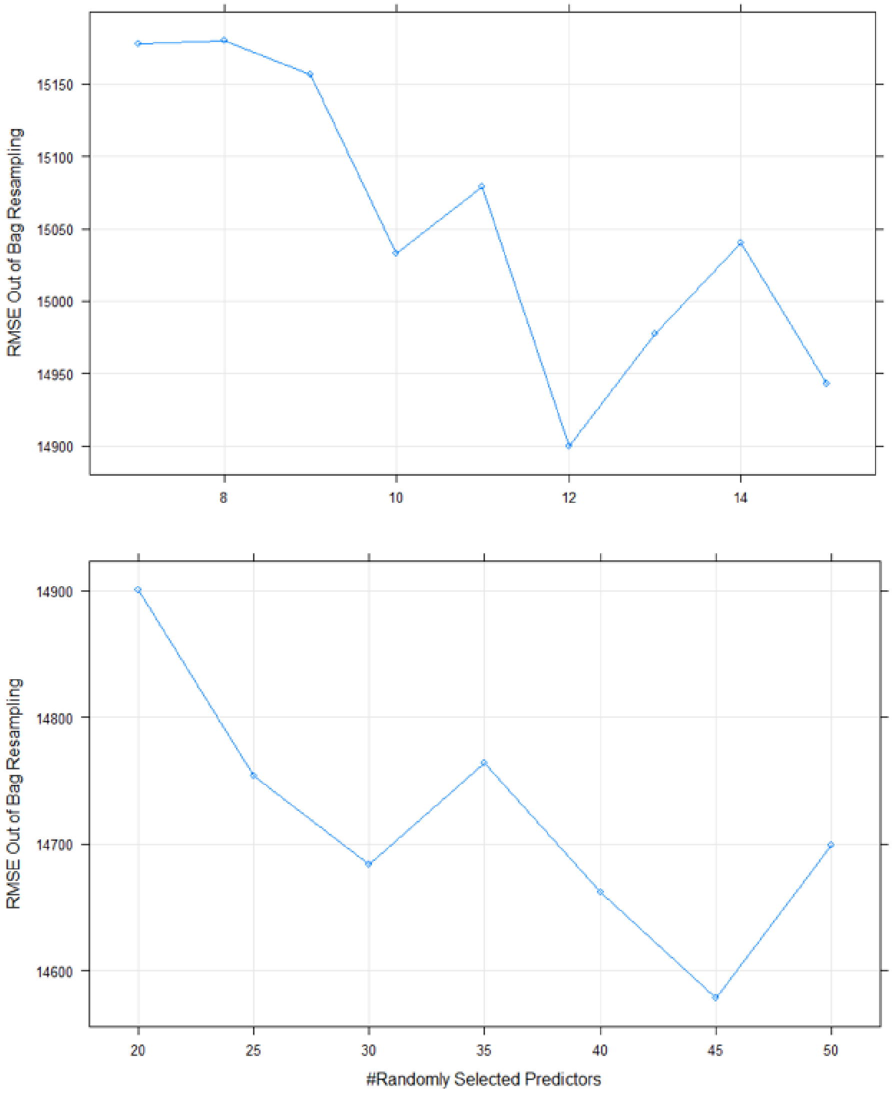
Table 3: Results of random search for values of mrty between 3 and 60

No. of variables per split (mrty)	R-squared
4	15.34
8	20.82
12	20.19
13	21.69
14	20.71
19	22.65
27	22.55
29	22.99
43	24.39
56	24.18

Note. These are the results of ten randomly chosen levels of mtry, based on out-of-bag data. A higher R-squared indicates better performance.

Based on the results from table 3, two grid searches were executed. A grid search is a comparison of predictive scores for all models with a range of parameters. The difference with random search is that the full grid is tested instead of several randomly chosen values. The first grid search consisted of a range from 7 mrty to 15 mtry, as this was close to the default value and close to the relatively well-performing 13 mrty. The second grid search was done in a range between 20 mrty and 50 mrty with intervals of 5, increasing the explored range. Figure 1 shows the two ranges of mtry plotted against the Root Mean Square Error (RMSE). A table with R-squared scores is provided in Appendix A. The two plots show that 45 mrty has the lowest RMSE. Other relatively well-performing levels of mtry are 12, 15 and 30 mtry.

Figure 1: Results for grid searches based on the random search results



Note. The results for two grid searches are plotted against root mean squared errors (RMSE). A lower value indicates better performance. The grid search was performed on out-of-bag data.

Guided by the results of the grid searches, four new models were constructed. All models had 100 trees, but model 1 could choose from 12 variables per node, model 2 could choose from 15 variables per node, model 3 could choose from 30 variables per node and model 4 could choose from 45 variables per node. These models were then tested using the split sample method.

The sample was randomly split into two pieces. One for training (8000 observations) and one for testing (2000 observations). The results for the five models on the testing data can be seen in table 4. It clearly shows that the tuned models were not able to significantly improve overall performance compared to the default model. Models 2 and 3 score worse than the default model for both R-squared and CPM. Model 4 scores slightly worse in R-squared and slightly better in CPM. However, model 1 performed slightly better than the default on both R-squared and CPM. Therefore, this model is chosen as the basis for the final model.

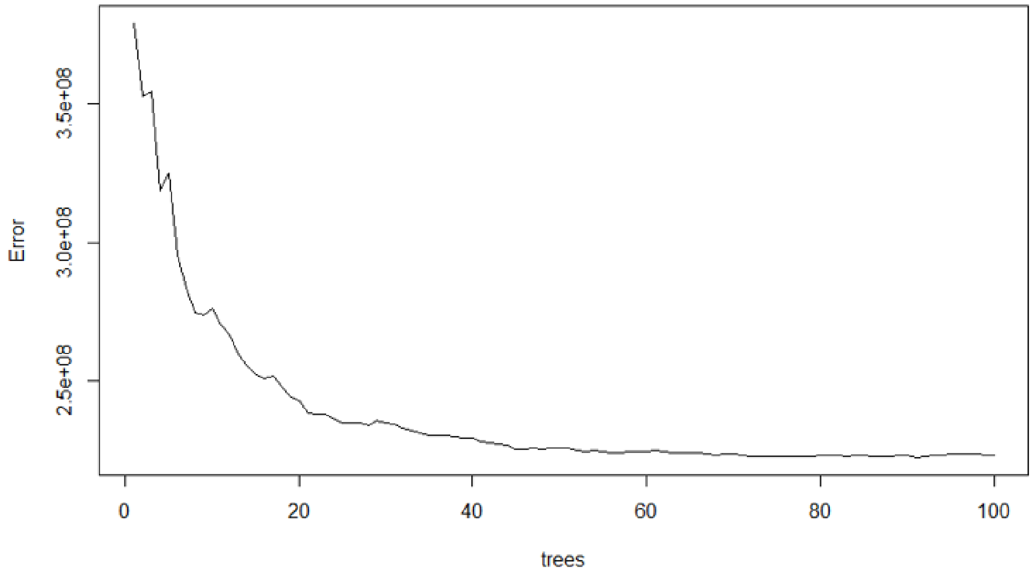
Table 4: Predictive performance of the default and tuning models

	R-squared	**CPM**
Default model (9 mrty)	0.0829	0.149
Tuned model 1 (12 mrty)	0.0874	0.150
Tuned model 2 (15 mtry)	0.0768	0.144
Tuned model 3 (30 mrty)	0.0732	0.141
Tuned model 4 (45 mtry)	0.0816	0.150

Tuning the number of trees

The following step was to establish the optimal number of trees. In principle, a larger number of trees will always improve performance, but due to diminishing marginal returns, the error rate will eventually stabilize, and additional trees will not meaningfully improve the model (Biau & Scornet, 2016). To check if model 1 has enough trees, the number of trees were plotted against the out-of-bag error rates. The results can be found in figure 2. Looking at this figure, it seems that the out-of-bag error rate starts to settle down around 50 trees. Therefore, it seems that model 3 has more than enough trees. To be on the safe side, the final model was run with 60 trees.

Figure 2: Number of trees plotted against the out-of-bag error rate



Note. The Y-axis shows the out-of-bag error rate. A lower value indicates a better performance. The X-axis shows the number of trees.

Final model

Considering the tuning results, the final model has 60 trees and 12 mrty. A plot of the number of trees against the out-of-bag error rate for the final model is provided in appendix C. The dependent variable is the residual healthcare spending for individual i in 2016. The independent variables are: (1) The full set of age groups for 2014, 2015 and 2016, (2) gender, (3) an indicator for hospitalization for 2014, 2015 and 2016, (4) the full set of dummy variables for PCGs for 2014, 2015 and 2016, and (5) healthcare spending in 2014 and 2015. The model was run on the full sample of 245,497 PCG-assigned individuals. The split sample method was used to test the model on unseen data. Train and test data were split with 30% of the data used for training and 70% used for testing.

Calculation over and under compensation

After the predictability of residual spending within PCGs was determined, the sample was divided into quintiles based on predicted residual spending. Next, the over and under compensation was determined per quintile. Based on those results, I was able to determine if any selection incentives exist within PCGs.

4.3.5. Supplementing the risk adjustment model with outlier risk sharing

In this step the Swiss risk adjustment model was supplemented with simulated outlier risk sharing. This consisted of an ex-post reimbursement for 100% of the costs above CHF 50,000 for every individual. To create the values for healthcare spending after risk sharing, the ‘normal’ healthcare spending in 2016 was taken as baseline, but all values above CHF 50,000 were set equal to CHF 50,000. Next, healthcare spending after risk sharing was used as the dependent variable in the replicated Swiss risk adjustment model. Residual spending after risk sharing was obtained by subtracting the predicted healthcare expenditures by the actual healthcare expenditures (maximized at CHF 50.000).

4.3.6. Comparing predictability in residual spending among PCGs before and after risk sharing

Next, residual spending after risk sharing was predicted by the final Random Forest model as discussed in paragraph 4.3.4. After the predictability in residual spending within the PCG group has been estimated with and without outlier risk sharing, the results for both situations were compared. This was done by dividing the sample in quintiles based on predicted residual spending. Finally, the over and under compensation per quintile was determined and compared. Based on the predictable profits and losses per quintile, I will be able to assess whether the introduction of risk sharing can reduce selection incentives within the PCG group.

Chapter 5: Results

5.1. Introduction

In this chapter, the results are discussed. First some descriptives are provided for the whole sample. Second, the Swiss risk adjustment model is replicated. Third, predictiveness of residual spending is estimated. Fourth, the effect of introducing risk sharing is discussed.

5.2. Descriptive statistics

Table 5 shows some key descriptive statistics for the full sample. After formatting the data, 805,531 observations were left. In 2016, the average healthcare expenditure for someone in the dataset was CHF 4,976 (in 2016, CHF 1 was roughly equal to EUR 0.91) The high standard deviation for healthcare expenditure (CHF 11,443) indicates that there is a high variance in the sample. The minimum healthcare expenditure was CHF 0. The maximum healthcare expenditure was CHF 1,405,606. The fact the average healthcare expenditure is CHF 4,976 within this range, indicates that the distribution of healthcare spending is skewed towards zero. Appendix B shows a graphical representation of this. Males make up 47.07% of the individuals in the sample and 8.58% of the individuals in the sample have been hospitalized in 2015 for at least three consecutive days. 30.48% of individuals have been assigned to at least one PCG in 2016, indicating pharmaceutical use in 2015.

Table 5: Descriptive statistics for the dataset of 2016.

Variable	N	Mean	Std. dev.	Min	Max
Healthcare expenditure	805,531	4,976	11,443	0	1,405,606
Male	805,531	47.07%	0.499	1	2
Hospitalized in 2015 [a]	805,531	8.58%	0.280	0	1
Assigned to at least 1 PCG [b]	805,531	30.48%	0.460	0	1

[a] *Indicates if someone has been hospitalized in 2015 for three or more consecutive days.*

[b] *Based on pharmaceutical use in 2015.*

In table 6, descriptive statistics have been split across individuals who are assigned to at least one PCG, and those who are not. The table shows that being assigned to at least one PCG increases the mean healthcare expenditure four-fold compared to those who are not assigned to a PCG. The percentage of people who were hospitalized for at least three consecutive days in 2015 for individuals who are not assigned to a PCG is 4.20%, while it is 18.60% for individuals who are assigned to at least one PCG. This indicates that individuals who are assigned to a PCG are more likely to have been hospitalized in the previous year. There are relatively fewer males in the group of individuals who are assigned to at least one PCG (41.90%), compared to individuals who are not assigned to any PCGs (49.30%). This indicates that females are more likely to be assigned to a PCG than males.

Table 6: Descriptive statistics of all individuals in 2016, split along those who are not assigned to any PCG, those who are assigned to at least one PCG and the total.

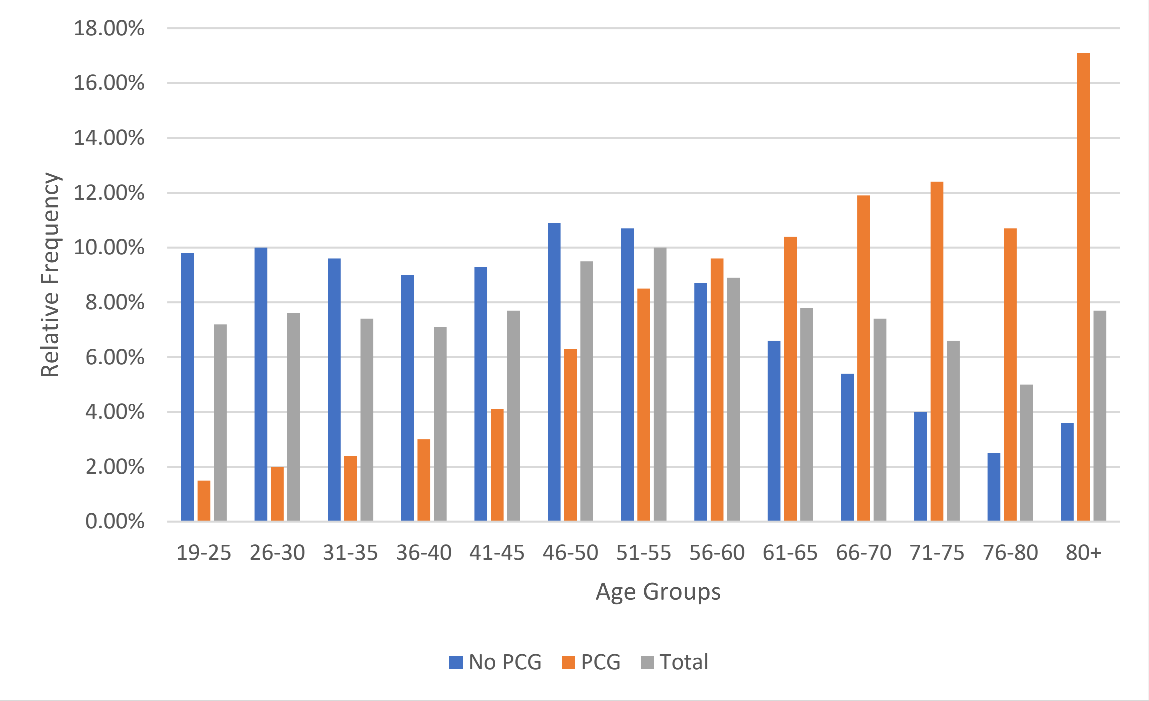
	Not assigned to a PCG	Assigned to PCG(s)	Total
N	560,034	245,497	805,531
Mean Healthcare expenditure	2,532	10,551	4,976
Hospitalized in 2015[a]	4.20%	18.60%	8.52%
Male	49.30%	41.90%	47.10%
Proportion assigned to PCG(s)[b]	0.00%	100.00%	30.05%

[a] *Indicates if someone has been hospitalized in 2015 for three or more consecutive days.*

[b] *Based on pharmaceutical use in 2015.*

Figure 3 shows the distribution of the sample across age groups, again split across the “PCG group” and the “no-PCG group”. For the PCG group, the most common age group is 81+ (17.1%), while only 1.5% of people with a PCG are aged between 19-25. For the no-PCG group, age distribution is roughly equal across age groups from 19-25 to 51-55 at circa 10.00%. However, older age groups are increasingly less represented. These results suggest that the probability of being assigned to at least one PCG increases with age.

Figure 3: Distribution of PCG individuals and non-PCG individuals across age groups



Based on the results shown in this paragraph, I conclude that 30.48% of the individuals in the sample are assigned to at least one PCG. Individuals who are assigned to at least one PCG differ from those who are not. They have higher healthcare expenditures, a higher chance to have been hospitalized in the previous year, and they are more likely to be female. In addition, those who are assigned to a PCG are likely to be older than those who are not assigned to a PCG.

5.3. Replicating the Swiss risk adjustment model

5.3.1. The estimated model

Table 7 shows a replication of the Swiss risk adjustment model. The reference category for age and gender are men aged between 19 and 25. All effects are statistically significant at a p-value of 0.001 except the effect for the male age group of 26-30 and PCG 15, which are statistically insignificant at a 0.05 level, ceteris paribus. Looking at the interaction terms between age and gender, it becomes clear that the predicted healthcare spending increases with age. Both for males and for females, higher age groups have higher coefficients. For example, a male aged 81+ is estimated to have CHF 5367 higher healthcare expenditure than a male aged 19-25. Females show the same. While being female aged 19-25 is estimated to increase healthcare spending by CHF 835 compared to a male aged 19-25, being a female aged 81+ is estimated to increase healthcare spending by CHF 5353 compared to a male aged 19-25, ceteris paribus.

What is also striking is that the coefficients suggest that females have a higher estimated healthcare expenditure than males in age groups 15-25 until 51-55. For older age groups, this is the other way around. One explanation could be that many females give birth during this period, which could raise their average healthcare expenditure. The fact that females become ‘cheaper’ than males from their fifties onwards might be explained by the fact that females have a higher life expectancy. Therefore, one could assume that females who are in the same age group as their male counterparts are more likely to be healthy and less likely to be in their last year of life.

The bottom variable in the left column (Hosp16) is an indication that an individual was hospitalized for at least three consecutive days in the previous year. If this is the case, estimated healthcare expenditure is estimated to increase by CHF 8339, ceteris paribus.

The right column shows the estimations for the different PCGs. Most PCGs are estimated to increase healthcare spending. However, there is significant variation. For example, PCG 13 is associated with an estimated increase of CHF 33,595, while PCGs 11 and 17 are associated with an estimated increase of less than CHF 1,000. Moreover, PCG 15 is not associated with an increase in healthcare spending at all. It is hard to pinpoint the source of these differences, as the medical nature of the PCGs is kept private. It could be the case that PCG 15 is a chronic illness that is associated with an otherwise healthy lifestyle or that the chronic illness is highly correlated with age, gender or hospitalization. Lastly, if an individual is not assigned to any PCG, this is associated with a decrease in healthcare spending of CHF 493.

The R-squared value for this model is 0.247. This means that 24.70% of the overall variance was explained by the model. This seems like a plausible result. Other risk adjustment schemes receive similar R-squared values. In addition, a large part of the demand for medical care is not predictable.

Table 7: Replicated Swiss risk adjustment model using healthcare spending in 2016 as the dependent variable.

Regressors	Coefficients	SE	Regressors	Coefficients	SE
F # 19-25	835***	(56.0)	pcg1_16	1570***	(108.4)
F # 26-30	1354***	(42.2)	pcg2_16	5790***	(217.7)
F # 31-35	1884***	(49.2)	pcg3_16	1784***	(99.1)
F # 36-40	1452***	(47.5)	pcg4_16	3377***	(169.7)
F # 41-45	1088***	(47.1)	pcg5_16	8479***	(295.8)
F # 46-50	1119***	(45.5)	pcg6_16	4300***	(68.6)
F # 51-55	1270***	(50.1)	pcg7_16	8005***	(348.6)
F # 56-60	1339***	(52.2)	pcg8_16	7071***	(258.2)
F # 61-65	1536***	(59.5)	pcg9_16	16247***	(994.0)
F # 66-70	2039***	(65.8)	pcg10_16	15193***	(325.8)
F # 71-75	2737***	(76.7)	pcg11_16	806***	(133.2)
F # 76-80	3395***	(93.9)	pcg12_16	19619***	(492.2)
F # 81+	5353***	(78.3)	pcg13_16	33595***	(1462.3)
M # 19-25	0	(.)	pcg14_16	2268***	(98.4)
M # 26-30	35	(52.0)	pcg15_16	-53	(78.1)
M # 31-35	202***	(51.6)	pcg16_16	1125***	(99.3)
M # 36-40	203***	(43.8)	pcg17_16	726***	(82.6)
M # 41-45	448***	(46.9)	pcg18_16	1084***	(190.9)
M # 46-50	612***	(51.3)	pcg19_16	1172***	(133.1)
M # 51-55	1039***	(62.8)	pcg20_16	3694***	(76.0)
M # 56-60	1495***	(60.0)	pcg21_16	4050***	(185.9)
M # 61-65	2168***	(74.7)	pcg22_16	3910***	(485.3)
M # 66-70	2776***	(83.1)	pcg23_16	4237***	(301.6)
M # 71-75	3823***	(104.4)	pcg24_16	5346***	(265.8)
M # 76-80	4470***	(118.7)	pcg25_16	5078***	(218.6)
M # 81+	5367***	(111.8)	pcg26_16	2934***	(271.9)
Hosp16	8339***	(78.7)	nopcg16	-493***	(66.4)
			Constant	1311***	(71.0)
Observations	805531				
R2	0.247				

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

5.3.2. Predictable profits and losses before and after risk adjustment

Table 8 shows the mean financial result in 2016 before and after risk adjustment for several subgroups that are explicitly accounted for in the risk adjustment model. Individuals who were hospitalized in the prior year for at least three consecutive days will on average be loss-making by CHF 12,410 before risk adjustment. For individuals who are assigned to at least one PCG, the average loss is CHF 5575. On the other hand, those who are not assigned to a PCG will be profitable for CHF 2444 on average before risk adjustment.

Because these three subgroups are explicitly part of the Swiss risk adjustment scheme as regressors, predictable profits and losses will be, on average, brought down to zero after the application of risk adjustment. This effectively eliminates selection incentives between these groups and others. However, there can still be considerable variation within groups that are explicitly part of the risk adjustment scheme. This variation can lead to predictable profits and losses. The next paragraph will discuss the predictability of residual spending within those who are assigned to PCG.

Table 8: Mean financial result in CHF in 2016

Subsample	Before risk adjustment	After risk adjustment
Not assigned to a PCG	CHF 2,444	CHF 0
Assigned to a PCG(s)	CHF -5,575	CHF 0
Hospitalized in the prior year	CHF -12,410	CHF 0

From the results in this section, it follows that according to the estimates of the Swiss risk adjustment scheme, older individuals are estimated to have higher healthcare costs than younger individuals. In addition, in the age groups 18-25 until 51-55, females are estimated to have higher healthcare expenditure than males. From age group 56-60 onwards, males are estimated to have higher healthcare expenditure. In addition, hospitalization and assignment to a PCG is associated with increased healthcare expenditure. However, large differences in estimated healthcare expenditure exist.

The Swiss risk adjustment scheme perfectly compensates for the average predictable profits and losses for groups that are explicitly part of the risk adjustment scheme. However, there can still be considerable variance within morbidity groups. This could potentially lead to selection incentives.

5.4. The predictability of residual spending variation within the PCG group

5.4.1. Performance of the Random Forest model

Table 9 shows performance indicators for the Random Forest model. It was tasked with predicting residual spending within the PCG group. The R-squared is 0.288, indicating that the Random Forest model was able to successfully predict 28.80% of the variance in the sample. Considering that the individuals in the sample were already explicitly part of the risk adjustment model, the Random Forest models seems to have considerable predictive power. Even after risk adjustment, this model was still able to predict almost a third of the variance in residual spending among PCG patients.

The CPM value is 0.190, which is significantly lower than the value for R-squared. This could result from the fact that CPM is less sensitive to observations with high error terms. If the Random Forest model was relatively good at predictions for individuals with very high healthcare costs, this could result in a higher R-squared score compared to the CPM score.

Table 9: Predictive performance of the Random Forest model

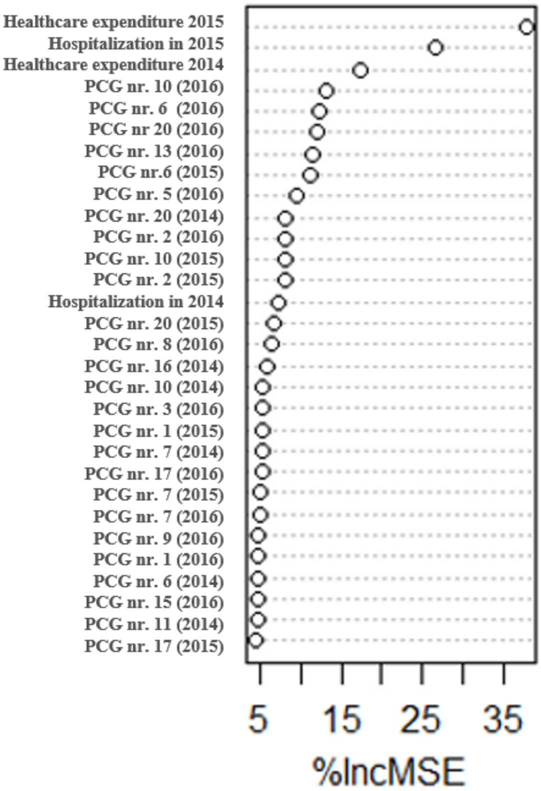
	R-squared	**CPM**
Final model	0.288	0.190

Figure 4 shows the relative importance of the variables that are used by the RF model. It measures the increase in the model's mean squared error (MSE) when each variable is randomly permuted while keeping other variables constant. The higher the %incMSE value for a variable, the greater its importance in predicting the target variable. Therefore, variables with larger %incMSE values are considered more influential in the RF model.

The most important variable was healthcare expenditure of an individual in 2015. One could say that this is not very surprising, as previous healthcare expenditure is not used in the Swiss risk adjustment model, while other variables like gender, age and PCGs are already part of the Swiss risk adjustment model. The fact that permuting healthcare spending in t -1 increased MSE by more than 35% indicates that healthcare spending in t -1 is a strong predictor for residual spending. Bruttoleist14 is the third most important variable according to this plot. This is likely to have the same reasons as bruttoleist15.

Perhaps more surprisingly, hospitalization in 2015 is the second most important variable. Permuting this variable leads to an increase in MSE of circa 27%. This might be called surprising because hospitalization is a variable that is part of the original risk adjustment model. Therefore, one might expect that a large part of the predictable variation caused by this variable is already excluded from residual spending. However, the importance of hospitalization might be increased if it has large interaction effects with other variables. A linear model like the Swiss risk adjustment model does not account for the existence of interaction between hospitalization and other variables like age, gender or PCGs. In contrast, a Random Forest model is able to find previously unknown and potentially complex interaction terms. For example, an interaction between hospitalization and being in a PCG. One could imagine that people with a chronic illness might have more chance for complications if they are hospitalized, leading to higher costs. If this is the case, running the model only on the PCG-group instead of the full sample might have increased the importance of hospitalization.

Figure 4: Variable importance plot



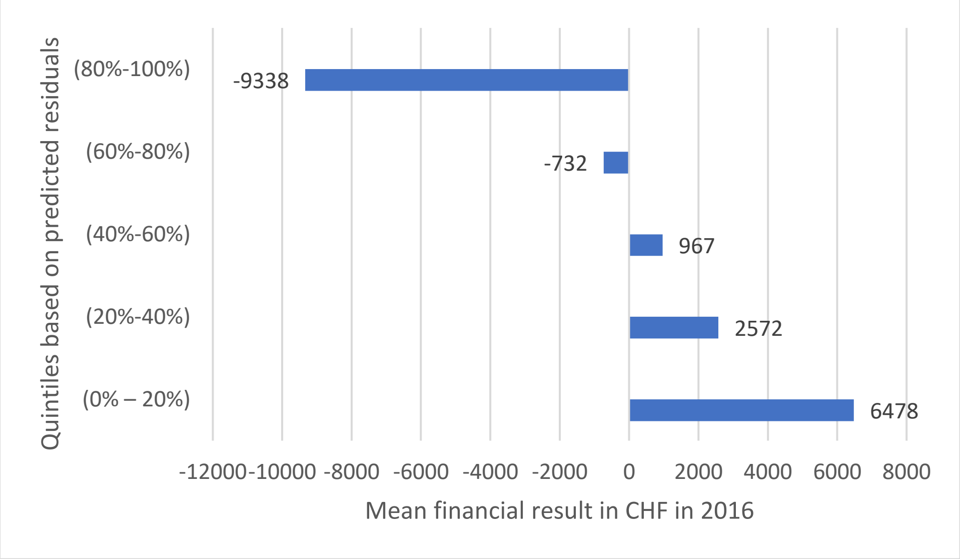
Note. The X-axis shows the percentage increase in Mean Squared Error after permuting a variable in the Random Forest model. A higher value indicates a bigger importance of the corresponding value. The Y-axis shows the most important variables in the model.

5.4.2. Selection incentives among the PCG group

While performance indicators like R-squared and CPM give a good idea of the predictability of residual spending in the PCG group, it does not tell us to what degree selection incentives exist within the sample. Figure 5 shows the predictable profits and losses for quintiles based on predicted residuals. The (0%- 20%) group is the 20% of individuals with the lowest predicted residual spending according to the Random Forest model. The (80%-100%) group is the 20% of individuals with the highest predicted residual spending according to the Random Forest Model.

From figure 5 it becomes clear that predictable profits and losses exist within the PCG group after risk adjustment. Those in the lowest quintile will on average be profitable for CHF 6478 per person per year. In stark contrast, the 20% with the highest predicted residual spending will on average be loss making for CHF 9338 per person per year.

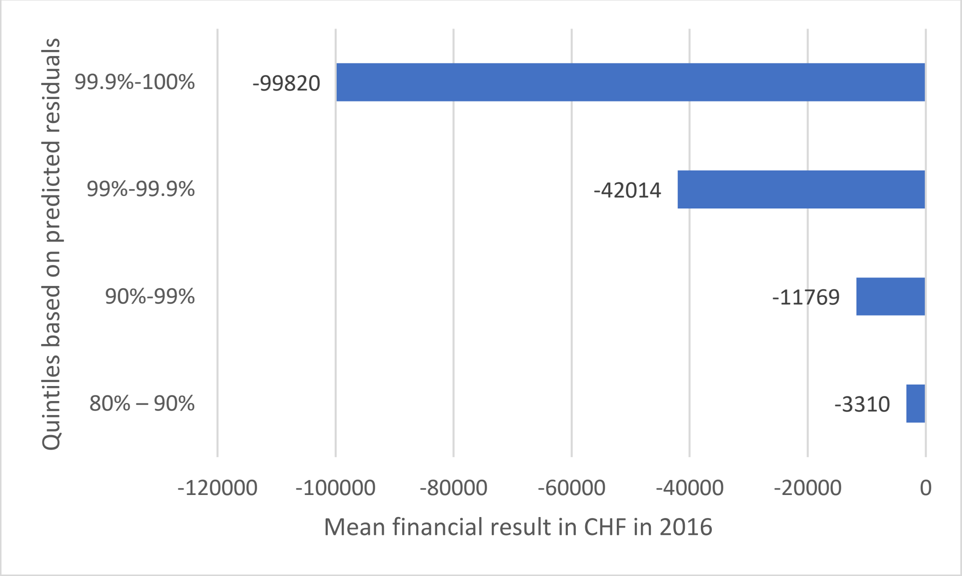
Figure 5: Financial results by quintiles based on predicted residuals



Note. The quintiles on the Y-axis are based on the predictions by the Random Forest model. Group (0%-20%) has the lowest predicted residual spending, group (80%-100) has the highest predictable spending. The mean financial result on the X-axis is derived by subtracting actual healthcare costs in 2016 from predicted healthcare costs (by the Swiss risk adjustment model). For all groups, the mean financial result is significantly different from zero at a 0.001 confidence level.

Figure 6 zooms in on the distribution of predictable losses in the top quintile (80%-100%). Again, subsamples are based on predicted residual spending. It becomes clear that within the top quintile, there is a very high variance between individuals. While the average predictable loss for group (80% - 90%) is CHF 3310, the top 1% have a predictable loss of CHF 42,014. The highest 0.1% has an even greater predicted loss of CHF 99,820 per person per year. The results from figures 5 and 6 indicate that after risk adjustment, there are still considerable predictable profits and losses within morbidity groups. In addition, predictable losses seem to be concentrated among those with the highest predicted residual spending. One method to potentially decrease predictable profits and losses among morbidity groups might be to introduce outlier risk sharing. The results of supplementing the Swiss risk adjustment scheme with risk sharing are discussed in the next paragraph.

Figure 6: Financial results for the top quintile based in predicted residuals



Note. The quintiles on the Y-axis are based on the predictions by the Random Forest model. Group (0%-20%) has the lowest predicted residual spending, group (80%-100) has the highest predictable spending. The mean financial result on the X-axis is derived by subtracting actual healthcare costs in 2016 from predicted healthcare costs (by the Swiss risk adjustment model). For all groups, the mean financial result is significantly different from zero at a 0.001 confidence level.

In this section, a Random Forest model was used to predict residual spending variation within the PCG group. The Random Forest model achieved an R-squared of 0.288, indicating that 28.80% of the variance in residual spending within the PCG group can be explained by the model. Het most important variables for the Random Forest model were healthcare expenditures from previous years and hospitalization in 2015. When looking at predictable profits and losses for groups based in their predicted residual spending, it becomes clear that considerable incentives for risk selection remain within the PCG group after risk adjustment. These incentives seem to concentrate at the individuals with the highest predicted residual spending.

5.5. The predictability of residual spending after introduction of risk sharing

For this paragraph, the Swiss risk adjustment scheme was supplemented with outlier risk sharing to test if the addition of risk sharing could decrease the predictability of residual spending within the PCG group. The chosen outlier risk sharing scheme applied to 100% of the healthcare costs above CHF 50.000 per person per year. Of these parameters were used, 6.16% of total the total healthcare expenditures would be reimbursed by a third party to the insurers. While this could possibly have a negative effect on incentives for cost consciousness, it might also decrease selection incentives within the PCG group.

To simulate outlier risk sharing, a new variable was created for healthcare expenditure that was maximized at CHF 50.000 (healthcare expenditure after risk sharing). Next, the Swiss risk adjustment scheme estimated new predicted healthcare expenditures for every individual, based on healthcare expenditure after risk sharing (predicted healthcare expenditure after risk sharing). These estimates are provided in appendix C. Based on the new predictions, residual spending after risk sharing was constructed by subtracting predicted healthcare expenditure after risk sharing from actual healthcare spending after risk sharing. Next, the Random Forest model was asked to predict residual spending after risk sharing.

Table 10 shows performance indicators for the Random Forest model after the introduction of risk sharing. The results for predicted residuals without risk sharing are included as a comparison.

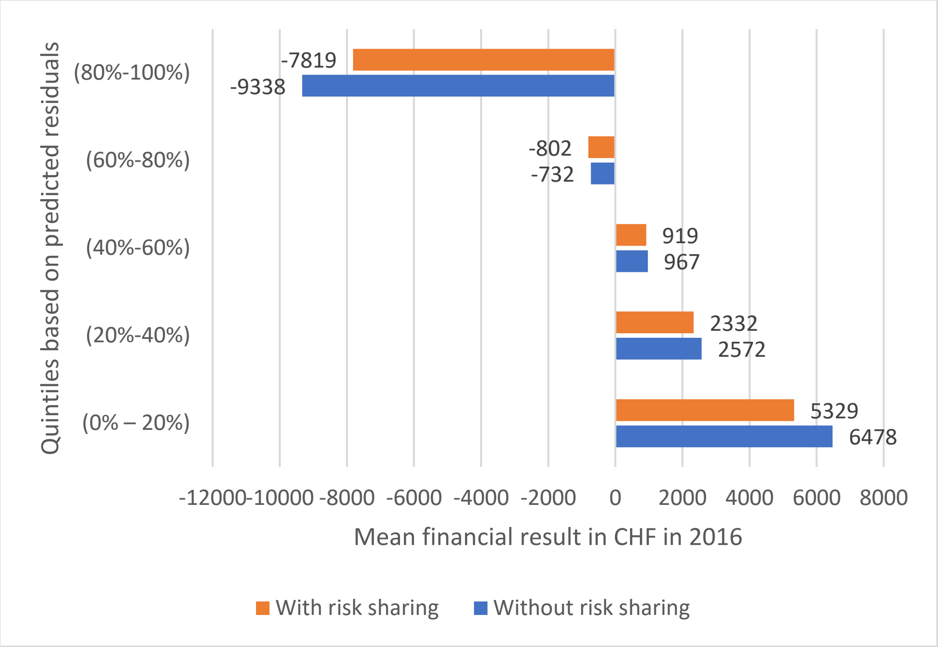
Table 10: Performance indicators for the Random Forest model without and with risk sharing

	R-squared	**CPM**
Without risk sharing	0.288	0.190
With risk sharing	0.280	0.197

While the performance indicators seem to suggest that the Random Forest model was able to predict residual spending after risk sharing to a similar degree as residual spending without risk sharing, it is not actually possible to directly compare these indicators. As can be seen in paragraph 3.4.1, both R-squared and CPM are partly based on the total variance in a sample. However, the introduction of risk sharing may inherently impact the total variance. Therefore, based on these performance indicators it is not possible to say if predictive residual spending has decreased due to the introduction of risk sharing. For a direct comparison, I will once again look at predictable profits and losses for subgroups.

Figure 7 shows predictable profits and losses per quintile based on predicted residual spending after risk sharing. A comparison is provided to the situation without risk sharing. It becomes clear that the addition of risk sharing decreases predictable profits and losses. The effect seems to be greatest at the highest and the lowest quintiles. For the lowest quintile, predictable profits decrease by CHF 1149 per person per year after the introduction of risk sharing. For the highest quintile, predictable losses decreased by CHF 1519 per person per year after the introduction of risk sharing.

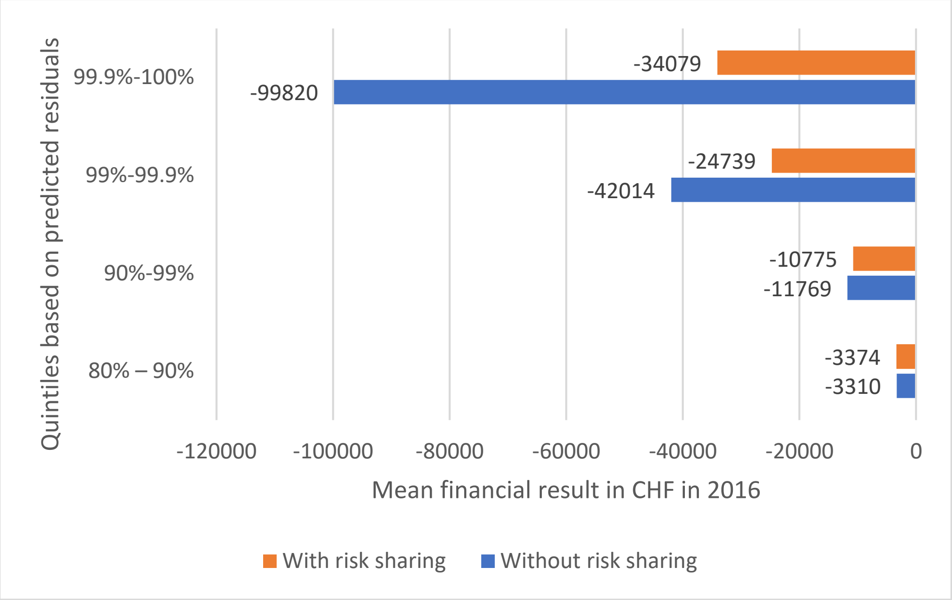
Figure 7: Predictable profits and losses per quintile based on predicted residual spending, both without risk sharing and with risk sharing



Note. The quintiles on the Y-axis are based on the predictions by the Random Forest model. Group (0%-20%) has the lowest predicted residual spending, group (80%-100) has the highest predictable spending. The mean financial result on the X-axis is derived by subtracting actual healthcare costs in 2016 from predicted healthcare costs (by the Swiss risk adjustment model). For all groups, the mean financial result is significantly different from zero at a 0.001 confidence level.

Figure 8 elaborates on the predictable losses in the upper quintile of figure 7. Again, a comparison to the situation without risk sharing is provided. If becomes clear that the decrease in predictable losses is concentrated at individuals with the highest predicted residual spending. For the (80%-90%) decile the difference is negligible. In fact, there is even a slight increase in predictable losses of CHF 64. However, for the 90%-99% group, predictable losses decrease by CHF 994. For the 99%-99.9% group predictable losses decrease by CHF 17,275 For the top 0.1%, introducing risk sharing decreases predictable losses by CHF 65,741. The larger differences for individuals with higher predicted healthcare expenditure might be explained by the fact that these individuals are the most likely to have healthcare expenditures higher than CHF 50,000. Therefore, their effective healthcare expenditure is most likely to be reduced by risk sharing.

Figure 8: Predictable profits and losses for the top quintile based on predicted residual spending, both without risk sharing and with risk sharing



Note. The quintiles on the Y-axis are based on the predictions by the Random Forest model. Group (0%-20%) has the lowest predicted residual spending, group (80%-100) has the highest predictable spending. The mean financial result on the X-axis is derived by subtracting actual healthcare costs in 2016 from predicted healthcare costs (by the Swiss risk adjustment model). For all groups, the mean financial result is significantly different from zero at a 0.001 confidence level.

In this paragraph, outlier risk sharing was added as a supplement to the Swiss risk adjustment model to analyse if this could reduce predictability in residual spending within the PCG group. I conclude that this is the case. Adding risk sharing can reduce selection incentives within the PCG groups. These effects are concentrated at individuals with the highest and the lowest predicted healthcare expenditure. When looking more closely at the group of individuals with the 20% highest predicted residual spending. The effect of adding risk sharing seems to be greatest for those with the highest predicted residual spending.

Chapter 6: Discussion & conclusion

6.1. Summary of key findings

The main research question of this paper was:

"To what extent do selection incentives exist within Pharmaceutical Cost Groups of the Swiss Risk Adjustment Scheme and to what extent can outlier risk sharing help reduce such selection incentives?"

The first key finding is that individuals who are assigned to at least one PCG have, on average, higher healthcare expenditure, are more likely to have been hospitalized in the previous year, are more likely to be female, and are more likely to be older than those who are not assigned to a PCG. Without risk adjustment, these individuals would be predictably loss-making for insurers. Currently, PCGs are explicitly used in the Swiss risk adjustment scheme. This results in the fact that the increased healthcare expenditure for individuals who are assigned to a PCG are, on average, perfectly compensated. However, within the PCG group there can still be variation.

The second key finding is that considerable selection incentives exist within the PCG group. By using a Random Forest model, I was able to estimate that 28.80% of the variance in residual spending within the PCG group is predictable. Based on the predictions of this model, I identified considerable selection incentives within the PCG group. The selection incentives seem to concentrate at individuals with the highest and the lowest predicted residual spending. The 20% of individuals with the lowest predicted residual spending were on average profitable for CHF 6478 per person per year. The 20% of individuals with the highest predicted residual spending were on average loss making for CHF 9338 per person per year.

The third key finding is that outlier risk sharing can decrease selection incentives within the PCG group. The Swiss risk adjustment model was supplemented with outlier risk sharing for 100% of the costs above CHF 50.000 per person per year. Based on the predictions of the Random Forest model, I found that the addition of outlier risk sharing decreases selection incentives within the PCG group. Predictable profits and losses decreased for most individuals, but the effect was largest for the most profitable and the most loss-making individuals. The predictable profitability of the 20% of individuals with the lowest predicted residual spending was reduced by CHF 1149 on average per person per year. The 20% of individuals with the highest predicted residual spending were on average loss making for CHF 1519 on average per person per year.

6.2. Contribution to the literature

The key scientific novelty in this paper is that selection incentives were investigated *within* the PCG group. Most literature focusses on differences *between* groups that can result in selection incentives. I build on this research by focussing on differences *within* a group. More precisely, I focus on individuals who are assigned to a PCG. Through this new approach, I was able to quantify how much predictable profits and losses remain within a the PCG group after compensation by the Swiss risk adjustment Scheme.

Secondly, by making use of a machine learning model, I added to the growing amount of research that utilizes machine learning in order to improve risk adjustment. I was able to show that a Random Forest model can predict 28.80% of residual spending variation within the PCG group, thereby adding to evidence that machine learning is a useful tool for predicting health expenditure.

6.3. Limitations

The first set of limitations concern the representativeness of the dataset. Individuals who were not available in all three years (2014 – 2016) were removed from the dataset because the Random Forest model does not allow for missing values. This means that individuals who changed their insurer during this period were excluded. If people who change insurer are on average healthier than those who do not change insurer, this could negatively impact the representativeness of the sample. Due to excluding switchers, the dataset might have higher average healthcare spending then in reality. In addition, healthcare spending of individuals who are similar to switchers might be overestimated.

Similarly, people who died in 2014 and 2015 were excluded from the dataset. If individuals in their last year of live have higher than average healthcare costs, excluding them might negatively impact the representativeness of the sample. By excluding people who died, the dataset might have lower average healthcare spending then in reality. In addition, healthcare spending for individuals who are similar to those who died might be underestimated.

Lastly, individuals who died in 2016 were kept in the sample, but they were not weighted for the time they were available during the year. This might underestimate the costs of the last year of life. Assuming that the people who died are among the most expensive individuals in 2016, not weighting them might not only underestimate the cost of the last year of life and underestimate costs for similar individuals, but it might also affect the skewedness of the sample.

If the healthcare costs of some groups are over or underestimated, the estimates for predictable profits and losses could also be over or underestimated for those groups. In this paper, both switchers and people in their last year of life were excluded, because the two groups can lead to opposing biases, they might have counter-acted each other to some degree. However, the effects of excluding these individuals were not tested in this paper. In future research these issues can be solved by using weighted and imputed data.

The second set of limitations concern the replication of the Swiss risk adjustment model. First, the original risk adjustment model is estimated individually for all the 26 cantons. I decided to calculate the model for the whole of Switzerland in order to maximize the number of observations. However, if the level or the distribution of healthcare spending differs per canton, the replicated model might not be faithful to the regional risk adjustment models. Second, several age groups were combined into the 81+ age group for both genders due to privacy regulations. This means that for individual aged above 81+ I could make less accurate estimations than the original model. These issues might be solved in future research by using larger datasets.

The third set of limitations concern the Random Forest model. Due to a lack of computational power, the training dataset was only 30% of the full sample. Normally training data consists of up to 80% of the full sample. In addition, the tuning dataset consisted of a subsample of 10,000 observations instead of the full dataset. It is possible that this decreased the predictive power of the model. With access to more computational power, larger training and tuning dataset can be used. This way, future research might be able to increase the predictive power of machine learning models.

If machine learning models are more capable of predicting residual spending when they have access to larger computational capacity, this could mean that future models might be able to find more predictable variation in residual spending. As a result, there might be more predictable profits and losses within the PCG group than predicted by the Random Forest model. This means that the predictable profits and losses that were shown in this paper might show the lower bound of the actual predictable profits and losses.

An additional limitation of the Random Forest model was that it did not predict residual spending for every individual PCG, but for all individuals who were assigned to at least one PCG. This was done because not all PCGs had enough observations for individual analysis and because the PCGs were anonymous. Therefore, useful interpretation would have been impossible. The disadvantage of this is that actual predictable profits and losses per PCG might differ from the overall estimates.

The final limitation concerns external validity. This research is based on data from a single health insurer in Switzerland. If this insurer has a specific consumer case-mix, the data might not be representative of entire Swiss population. On a broader scope one could question if the Swiss population is representative for other countries. There might be considerable differences in health, wealth and lifestyle between countries. In addition, the Swiss risk adjustment model differs from models used in other countries. All of this implies that the results in this paper cannot be generalized to other countries without thought. If the Swiss health insurer attracts specific individuals, the results might not even be generalizable to the rest of Switzerland.

6.4. Recommendations for further research

First, I would recommend future research to improve upon the findings in this paper by picking some 'low-hanging fruits'. Using imputed and weighted data will ensure the representativeness of the dataset and increasing computational power and sample size can improve the precision of the results.

If future research has access to larger and more externally representative datasets and greater computational capacity, the Random Forest model could be tuned and trained on larger datasets. This will likely improve the predictive power of the Random Forest model. This will allow future research to better establish to what extent selection incentives exist within the PCG group. In addition, future research could focus on more specific subgroups. For example, individual PCGs in specific cantons. This would give the results more real-life relevance.

Second, I would recommend future research to look beyond the PCG group and investigate selection incentives within other groups that are part of risk adjustment models. The variance and distribution of healthcare expenditure within regressor groups will most likely differ. New research could focus on identifying selection incentives in other regressor groups. For the Swiss risk adjustment model, this could be individuals who were hospitalized in the previous year or specific PCGs instead of the full PCG group. In other countries, many risk adjustment schemes use different regressor groups based on previous healthcare expenditure or like diagnostic cost groups (DCG). There can be selection incentives within all these groups as well.

Fourth, I would recommend future research to further investigate how to reduce selection incentives within regressor groups. This paper focussed on outlier risk sharing as a possible solution, but many alternatives exist. Other risk sharing regimes might yield different results. One possibility would be to investigate the effects of *high-risk pooling*. This is (outlier) risk sharing, but exclusively for a specific group of high risks like the PCG group. This might be more effective or more efficient in reducing selection incentives within the PCG group than the outlier risk sharing scheme used in this paper.

A different approach to lowering selection incentives within regressor groups is changing the risk adjustment model itself. One possibility that could be explored in future research is adding PCG tiers. One PCG could be split up into several tiers. A higher tier would have a higher threshold for drug use. This would reduce the size of the PCG group and, in theory, make it possible to more accurately predict individuals' healthcare expenditures. The same could be done for the hospitalization regressor. This could be split up into DCGs. Other countries like the Netherland have already implemented these measures. In principle this could yield the same results and PCG tiers.

6.5. Policy implications

The key finding in this paper is that selection incentives exist within morbidity groups and that these incentives are concentrated at the individuals with the highest and the lowest predicted residual spending respectively.

The main implication is that considerable incentives for risk selection exist in Switzerland, despite the risk adjustment scheme. In the theoretical framework I discussed the negative effects that risk selection can have on the functioning of healthcare systems. An insurer could try to increase its profits by dissuading unprofitable individuals from enrolling. To do this, an insurer could purposefully contract bad quality care if it concerns care that is attractive to unprofitable individuals. In this case, premiums will not only reflect differences in quality and efficiency of care, but also differences in health of the participants of an insurance. Policy makers should be aware of the issue of risk selection within the Swiss healthcare system, and they should endeavour to minimize this risk.

One option that could decrease selection incentives is outlier risk sharing. This paper investigated how outlier risk sharing could reduce the existence of risk selection incentives within the PCG group. The second key policy implication is that risk sharing can reduce selection incentives within morbidity groups. The effects of risk sharing seem to concentrate at those individuals with the highest and lowest predicted residual spending respectively. However, the outlier risk sharing scheme that was used in this paper, would mean that 6.16% of healthcare expenditure would be reimbursed to the insurers. This can potentially decrease their incentive for cost-conscious behaviour. After all, they do not bear the full costs of inefficiency.

This implies that outlier risk sharing can help to reduce selection incentives within regressor groups, but does so at a cost, in the shape of a reduction in cost-consciousness incentives. Policy makers should consider the positive effects that outlier risk sharing can have on selection incentives and might consider outlier risk sharing as a supplement to risk adjustment. However, they should be aware of the trade-off between risk selection and cost-consciousness incentives.

Next to outlier risk sharing, other options exist that might decrease selection incentives within regressor groups. While these alternatives were not investigated in this paper, the theory and results might have given some indications that could be helpful.

In the theorical framework I identified four types of risk sharing. Outlier risk sharing is only one of them. In theory, any type of risk sharing could be considered as a supplement to the risk adjustment model. In practice, the Swiss cantons already provide proportional risk sharing for hospitalization. One option that might be considered is *high-risk pooling*. This is risk sharing for an exclusive group of individuals. Let's consider that all individuals in the PCG group are assigned to the high-risk pool. In this scenario, an insurer would only be reimbursed if someone in the high-risk group meets the requirements for risk sharing. One potential advantage over 'normal' outlier risk sharing is that it might be a more affordable way of reducing selection incentive within specific groups. If the objective is to reduce selection incentives within specific groups, for example the PCG group, policy makers might consider *high-risk pooling* as a supplement to risk adjustment.

In addition to risk sharing, alternative risk adjustment models might also decrease selection incentives within regressor groups. Again, these alternatives were not investigated in this paper. However, in the theoretical framework and the results of this paper, some alternatives were discussed.

In the theoretical framework I discussed the fact that risk adjustments models have been improving in many countries by adding PCGs, DCGs and indicators of historical healthcare expenditure. The Swiss risk adjustment model does not use of all these regressors. Currently it has PCGs, a general indicator for hospitalization in the previous year, and no indicator for prior expenditure. Policy makers might consider expanding the risk adjustment model by splitting the hospitalization indicator into DCGs, and potentially adding indicators for prior healthcare expenditure. An argument in favour of this might be that in the results section of this paper, it became clear that prior hospitalization and healthcare expenditures were the most important predictors of residual spending. This could indicate that adding DGCs or prior healthcare expenditure indicators could decrease predictable profits and losses.

Policy makers should consider though, that DCGs and prior cost indicators might suffer from a trade-off between risk selection and cost-consciousness incentives similar to risk sharing. For example, if a treatment in a hospital is done inefficiently, this could theoretically increase the value for the corresponding DCG is the risk adjustment model for the next year. This way, incentives for cost-consciousness might decrease.

Another, more radical, alternative is to make use of machine learning in risk adjustment. As discussed in the theoretical framework, machine learning has its drawbacks. The predictions are not based on a transparent set of regressors. This makes some machine learning models difficult to interpret. However, if transparency is not one of the objectives, policy makers might consider machine learning as a viable alternative. Literature shows that machine learning consistently outperforms linear regressions. In this paper, the Random Forest model was able to predict 28.80% of residual spending. These levels of explanatory power are not uncommon among linear risk adjustment schemes. However, the Random Forest model predicted spending after risk adjustment was already applied. Taking this into account, machine learning might be the next big step forward in risk adjustment.

References

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *The American Economic Review*, pp. 141-149.

Ash, S. A., McCall, N., Fonda, J., Hanchate, A., & Speckman, J. (2005). *Risk Assessment of Military Populations to Predict Health Care Cost and Utilization.* Falls Church: Center for Health Care Management Studies.

Beck, K., Kauer, L., McGuire, T. G., & Schmid, C. P. (2020). Improving risk-equalization in Switzerland: Effects of alternative. *Health Policy*, 1363-1367.

Biau, G., & Scornet, E. (2016). A random forest guided tour. *TEST*, pp. 197-227.

Breiman, L. (2001). Random Forests. *Machine learning*, pp. 5-32.

Buchmueller, T., & DiNardo, J. (2002). Did Community Rating Induce an Adverse Selection Death Spiral? Evidence from New York, Pennsylvania, and Connecticut. *American Economic Review*, 280-294.

Chao, Y.-S., Wu, C.-J., & Chen, T.-S. (2014). Risk adjustment and observation time: comparison. *Health Information Science and Systems*.

Diamond, P. (1992, november). Organizing the health insurence market. *Journal of the Econometric Society*, pp. 1233-1254.

Ellis, R. P., Martins, B., & Rose, S. (2017). Risk Adjustment for Health Plan Payment. In G. T. McGuire, & R. van Kleef, *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice.* Elsevier Press.

Enthoven, A. C. (1978). Consumer-choice health plan. *The New England Journal of Medicine*, pp. 650-658.

Enthoven, A. C. (1993, March 1). The History and Principles of Managed Competition. *Health Affairs*, pp. 24-48.

Faraway, J. J. (2016). Does data splitting improve prediction? *Statistics and Computing*, 49-60.

Hastie, T., Friedman, J., & Tibshirani, R. (2009). *The Elements of Statistical Learning.* New York: Springer.

Kauer, L., McGuire, T. G., & Beck, K. (2020). Extreme under and overcompensation in morbidity-based health plan payments: The case of Switzerland. *Health policy*, 61-68.

Layton, T. J., & Ellis, R. P. (2018). Evaluating the Performance of Health Plan Payment Systems. In R. C. van Kleef, & T. G. McGuire, *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets.* (pp. 133-167). Elservier.

Mazumdar, M. L. (2020). Comparison of statistical and machine learning models for healthcare costs data: a simulation study motivated by Oncology Care Model (OCM) data. *BMC Health Sevices Research*.

McGuire, G. T., & van Kleef, R. (2018). Risk Sharing. In G. T. McGuire, & R. van Kleef, *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets.* (pp. 105-131). Elsevier.

McGuire, T. G., & van Kleef, R. C. (2018). Risk Sharing. In T. G. McGuire, & R. C. van Kleef, *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets.* (pp. 105-131). Elsevier.

Newhouse, J. P. (1986). Rate adjusters for medicare under capitation. *Health Care Financing Review*, pp. 45-55.

Nguyen, Q. H., Ly, H.-B., Ho, L. S., Al-Ansari, N., Van Le, H., Prakash, I., . . . Pam, B. T. (2021). Influence of Data Splitting on Performance of Machine Learning Models in Prediction of Shear Strength of Soil. *Mathematical Problems in Engineering*.

Powers, C. A., Meyer, C. M., Roebuck, C., & Vaziri, B. (2005). Predictive Modeling of Total Healthcare Costs. *Medical care*, 1065-1072.

Schmid, C. P., Beck, K., & Kauer, L. (2018). Health Plan Payment in Switzerland. In T. G. McGuire, & R. van Kleef, *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice.* (pp. 453-489). Elsevier Press.

van de Ven, W. P., Beck, K., Buchner, F., Schokkaert, E., Schut, E. F., Shmueli, A., & Wasem, J. (2013). Preconditions for efficiency and affordability in competitive healthcare markets: Are they fulfilled in Belgium, Germany, Israel, the Netherlands and Switzerland? *Health Policy*, pp. 226-245.

van de Ven, W. P., Beck, K., van de Voorde, C., Wasem, J., & Zmora, I. (2007). Risk adjustment and risk selection in Europe: 6 years later. *Health Policy*, 162–179.

van Kleef, R. C., Eijkenaar, F., & van Vliet, R. C. (2020). Selection Incentives for Health Insurers in Sophiticated Risk Adjustment. *Medical Care Research and Review*, pp. 584-595.

van Kleef, R. C., McGuire, T. G., van Vliet, R. C., & van de Ven, W. P. (2017, december 10). Improving risk equalization with constrained regression. *European Journal of health Economics*, pp. 1137-1156.

van Kleef, R., Reuser, M., Stam, P. J., & van de Ven, W. P. (2022, November). Positive and negative effects of risk equalization and risk sharing in regulated competitive health insurence markets. *EsCHER*.

van Veen, S., van Kleef, R., van de Ven, W., & van Vliet, R. (2015). Is There One Measure-of-Fit That Fits All? A Taxonomy and Review of Measures-of-Fit for Risk-Equalization Models. *Medical Care Research and Review*, pp. 220-243.

van Veen, S., van Kleef, R., van de Ven, W., & van Vliet, R. (2018). Exploring the predictive power of interaction terms in a sophisticated risk equalization model using regression trees. *Health Economics*, pp. e1-e12.

Vimont, A., Leleu, H., & Durand-Zaleski, I. (2022). Machine learning versus regression modelling in predicting individual. *The European Journal of Health Economics*, 211-223.

Ying, X. (2019). An overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*.

Appendix

Appendix A: Results for the two grid searches for the optimal value of mtry

Grid search 1: Close to the default value

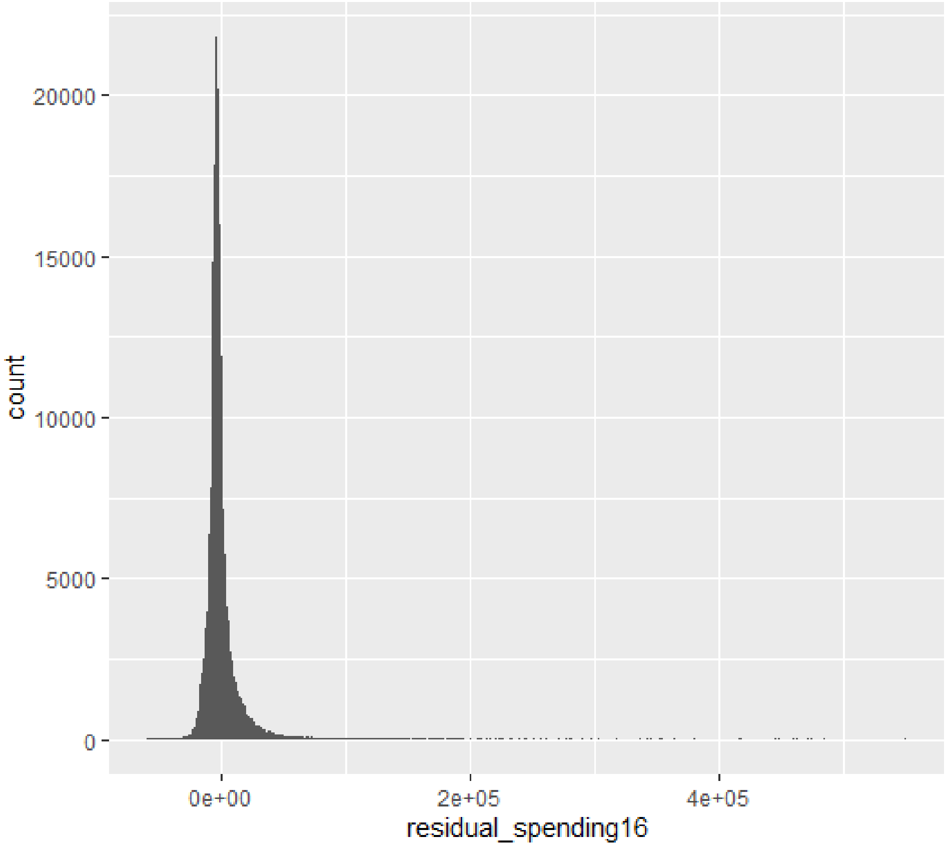
No. of variables per split (mrty)	**R-squared**
7	18.59
8	18.56
9	18.81
10	20.14
11	19.64
12	21.55
13	20.72
14	20.06
15	21.09

Grid search 2: wider range of mrty, based on random search results.

No. of variables per split (mrty)	**R-squared**
20	21.54
25	23.08
30	23.80
35	22.97
40	24.03
45	24.89
50	23.65

Note. This table contains two grid searches. Grid search one 1 was conducted near the default mtry value, grid search two was based on the random search results. The left column shows the level of mtry used in the Random Forest model. The right column shows the R-squared values.

Appendix B: Histogram containing the distribution of residual spending across the sample.



Note. The Y-axis shows the frequency of a particular level of residual spending, the X-axis shows residual spending. The median level of residual spending seems to be close to zero. In addition, the distribution seems the be skewed towards zero.

Appendix C: Comparison between the coefficients of the Swiss risk adjustment model before and after risk sharing.

	Before risk sharing	After risk sharing
F # 19-25	834.8***	780.2***
	(56.03)	(29.97)
F # 26-30	1353.6***	1354.3***
	(42.15)	(32.33)
F # 31-35	1883.9***	1856.3***
	(49.17)	(34.95)
F # 36-40	1451.9***	1435.3***
	(47.54)	(34.98)
F # 41-45	1087.6***	1068.4***
	(47.10)	(34.57)
F # 46-50	1118.6***	1115.5***
	(45.47)	(34.00)
F # 51-55	1269.8***	1262.1***
	(50.12)	(35.77)
F # 56-60	1338.6***	1335.6***
	(52.22)	(39.17)
F # 61-65	1536.1***	1509.6***
	(59.50)	(44.24)
F # 66-70	2039.1***	2005.7***
	(65.81)	(49.25)
F # 71-75	2737.3***	2731.0***
	(76.72)	(57.54)
F # 76-80	3395.4***	3439.0***
	(93.93)	(69.34)
F # 81+	5353.2***	5793.6***
	(78.25)	(64.51)
M # 19-25	0	0
	(.)	(.)
M # 26-30	34.58	-11.86
	(51.96)	(27.71)

M # 31-35	201.9^{***}	120.5^{***}
	(51.59)	(29.84)
M # 36-40	203.4^{***}	184.3^{***}
	(43.76)	(31.39)
M # 41-45	448.2^{***}	416.5^{***}
	(46.85)	(33.48)
M # 46-50	611.8^{***}	524.7^{***}
	(51.29)	(32.88)
M # 51-55	1038.6^{***}	889.4^{***}
	(62.79)	(35.52)
M # 56-60	1494.9^{***}	1324.9^{***}
	(59.96)	(41.04)
M # 61-65	2168.3^{***}	1863.0^{***}
	(74.65)	(48.51)
M # 66-70	2776.2^{***}	2463.0^{***}
	(83.14)	(55.74)
M # 71-75	3822.7^{***}	3373.2^{***}
	(104.4)	(67.46)
M # 76-80	4469.8^{***}	4146.6^{***}
	(118.7)	(84.72)
M # 81+	5367.0^{***}	5377.4^{***}
	(111.8)	(85.28)
hosp16	8338.6^{***}	7280.8^{***}
	(78.74)	(53.43)
pcg1_16	1570.3^{***}	1368.1^{***}
	(108.4)	(74.02)
pcg2_16	5790.4^{***}	5045.8^{***}
	(217.7)	(148.7)
pcg3_16	1784.0^{***}	1941.7^{***}
	(99.10)	(73.46)
pcg4_16	3377.2^{***}	2837.9^{***}
	(169.7)	(113.0)
pcg5_16	8479.3^{***}	8107.6^{***}
	(295.8)	(228.3)

pcg6_16	4299.5***	3628.4***
	(68.63)	(47.22)
pcg7_16	8005.0***	6219.8***
	(348.6)	(182.0)
pcg8_16	7070.9***	6543.6***
	(258.2)	(180.5)
pcg9_16	16246.5***	13245.2***
	(994.0)	(429.4)
pcg10_16	15193.3***	10489.4***
	(325.8)	(180.8)
pcg11_16	805.6***	855.0***
	(133.2)	(102.9)
pcg12_16	19619.2***	17877.1***
	(492.2)	(237.4)
pcg13_16	33595.3***	15220.6***
	(1462.3)	(535.0)
pcg14_16	2267.6***	2213.0***
	(98.43)	(71.78)
pcg15_16	-53.08	12.66
	(78.06)	(55.14)
pcg16_16	1125.1***	1058.2***
	(99.27)	(74.11)
pcg17_16	726.1***	630.6***
	(82.55)	(60.35)
pcg18_16	1084.0***	1057.0***
	(190.9)	(136.6)
pcg19_16	1171.9***	1138.7***
	(133.1)	(102.8)
pcg20_16	3693.9***	3515.4***
	(75.95)	(54.20)
pcg21_16	4050.0***	4095.7***
	(185.9)	(151.3)
pcg22_16	3910.2***	3590.2***

	(485.3)	(362.9)
pcg23_16	4237.2***	4537.7***
	(301.6)	(269.2)
pcg24_16	5346.0***	4567.0***
	(265.8)	(173.6)
pcg25_16	5078.3***	4338.0***
	(218.6)	(138.9)
pcg26_16	2934.1***	2950.4***
	(271.9)	(226.6)
nopcg16	-492.7***	-663.0***
	(66.43)	(45.73)
Constant	1310.8***	1473.0***
	(71.03)	(49.44)
Observations	805531	805531
R^2	0.247	0.342

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$