Evaluating and Improving Risk Equalization with Health Survey Information

Anja A. Withagen-Koster

Evaluating and Improving Risk Equalization with Health Survey Information

Anja A. Withagen-Koster

Copyright 2022, Anja A. Withagen-Koster, the Netherlands

All rights reserved. No parts of this thesis may be reproduced, stored in a retrieval system or transmitted in any form or by any means, without the permission of the author or, when appropriate, of the publishers of the publications

Layout and printing by Optima Grafische Communicatie (www.ogc.nl)

ISBN 978-94-6361-741-3

Evaluating and Improving Risk Equalization with Health Survey Information

Evalueren en verbeteren van de risicoverevening met informatie uit een gezondheidsenquête

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof. dr. A.L. Bredenoord

and in accordance with the decision of the Doctorate Board.
The public defence shall be held on

Wednesday the 23rd of November 2022 at 10.30 hrs

by

Anja Adriana Koster
born in Dordrecht, Netherlands.

Erasmus University Rotterdam

Doctoral Committee:

Promotor:	Prof. dr. F.T. Schut
Other members:	Prof. dr. E. Schokkaert Prof. dr. J.D. de Jong dr. ir. R.C.M.H. Douven
Copromotors:	dr. R.C. van Kleef dr. F. Eijkenaar

CONTENTS

LIST OF PUBLICATIONS AND SUBMISSIONS

Chapter 2, 3, 4, 5 and 6 are based on the following articles:

Chapter 2

Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. (2018). Examining unpriced risk heterogeneity in the Dutch health insurance market. European Journal of Health Economics. https://doi.org/10.1007/s10198-018-0979-x

Chapter 3

Withagen-Koster, A.A., van Kleef, R.C., Eijkenaar, F. Predictable profits and losses in a health insurance market with risk equalization: a multiple-contract period perspective. Submitted

Chapter 4

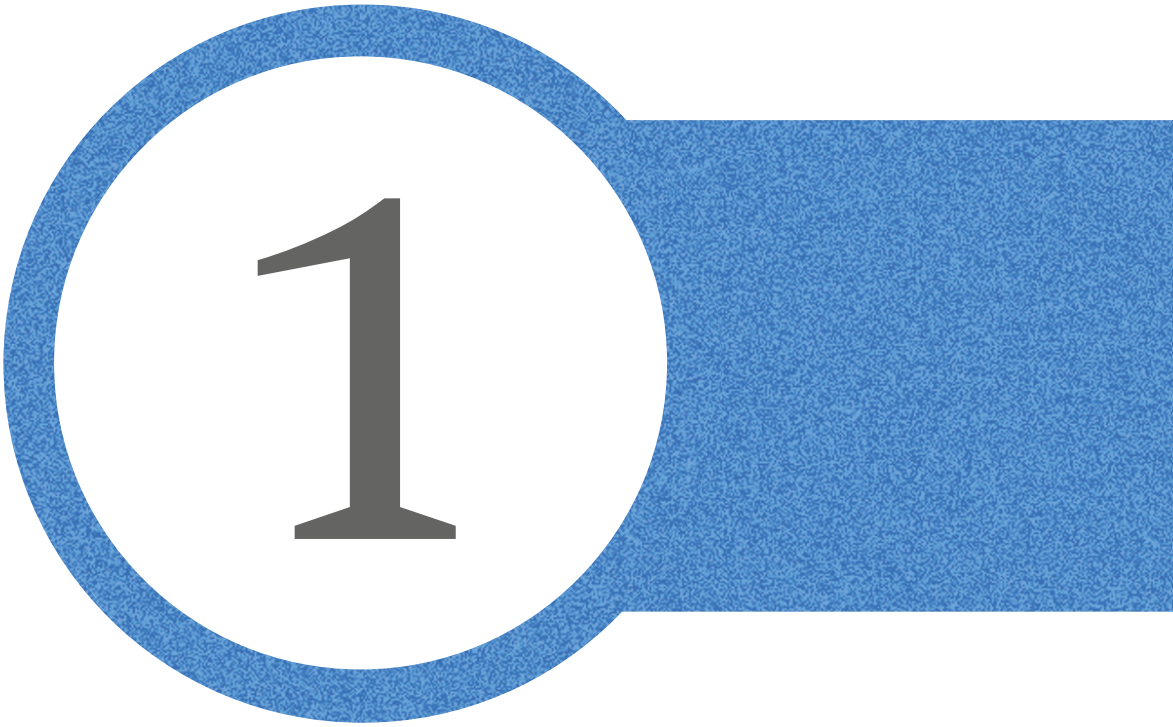
Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. (2022). Selection incentives in the Dutch basic health insurance: to what extent does end-of-life spending contribute to predictable profits and losses for selective groups? Medical Care Research and Review. https://doi.org/10.1177/10775587221099731

Chapter 5

Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. (2020). Incorporating self-reported health measures in risk equalization through constrained regression. European Journal of Health Economics. https://doi.org/10.1007/s10198-019-01146-y

Chapter 6

Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. An innovative method to mitigate risk selection incentives in health insurance markets with sophisticated risk equalization: high-risk pooling based on health survey information. Submitted


1

Introduction

1.1 BACKGROUND

Visiting a doctor when feeling ill or collecting medication at the pharmacy is self-evident for most people living in developed countries. Even having surgery when necessary is often considered normal. Receiving all required health care without going bankrupt is possible due to health insurance. Health insurance makes otherwise unaffordable health care affordable to the population. However, on a *free competitive* health insurance market, health insurance would not be affordable for every individual as the premium would reflect the financial risk (i.e., expected spending) of an individual. The reason for this is that in a free competitive market, a health insurer must break even on every insurance contract. After all, if a health insurer does not adjust the premium to the financial risk, low-risk individuals will switch to another insurer that offers a similar health plan in exchange for a lower premium. The first insurer is then left with only relatively high-risk individuals and forced to increase his premium. This way, premiums on a *free competitive* health insurance market will reflect the risk of an individual (i.e., this is called the equivalence principle) (Van de Ven & Ellis 2000). For high-risk individuals (e.g., individuals suffering from chronic illnesses) health insurance might then become unaffordable. This is one of the reasons why many countries have reformed their health insurance system towards a system of *regulated* competition.

In these countries, regulated competition is implemented to ensure efficiency, accessibility, and affordability of health insurance (Van de Ven et al. 2003; McGuire & Van Kleef 2018a). Within this system health insurers compete on price (premium) and (quality of) contracted provider network, which should lead to coverage of good-quality health care at a reasonable price. Regulation safeguards public goals such as affordability and accessibility of basic coverage. To that end, regulatory measures often include premium-rate restrictions to ensure affordability, especially for high-risk individuals. Premium-rate restrictions imply the transfer of money from low-risk individuals (for whom the premium exceeds their expected spending) to high-risk individuals (for whom the premium falls below their expected spending). In addition to this regulatory measure, there is often a mandate to buy health insurance. Without such a mandate, low-risk individuals might not buy health insurance as the premium they have to pay would be considered too high compared to their expected spending. Consequently, health insurers are confronted with incentives to attract low-risk individuals (i.e., healthy individuals) who are predictably profitable and to deter high-risk individuals (i.e., chronically ill) who are predictably unprofitable since premiums no longer reflect the risk of individuals (Van de Ven & Ellis 2000; McGuire & Van Kleef 2018b). To prevent health insurers from engaging in such

risk selection, regulations typically also include an open enrollment requirement (i.e., an obligation for insurers to accept all individuals who apply for a health plan), a standardized benefits package and risk equalization (McGuire & Van Kleef 2018b).

Risk equalization is an important instrument to mitigate risk-selection incentives in a competitive health insurance market with premium-rate restrictions. It compensates health insurers for predictable spending variation among individuals using a set of risk factors (i.e., risk adjusters). For example, risk-equalization payments are typically higher for the elderly than for the young, and higher for people with a chronic disease than for those without. Risk-equalization models currently used around the world range from relatively simple models including risk adjusters for age and gender (e.g., Israel) to sophisticated models that also include risk adjusters based on health (e.g., Belgium, Germany, Switzerland, US Marketplaces, US Medicare Advantage and the Netherlands). Nevertheless, none of these models compensate insurers completely for predictable spending variation. As a result, insurers still face predictable losses on chronically ill individuals and predictable profits on healthy individuals, implying that risk-selection incentives remain (Buchner et al. 2013; Ellis et al. 2013b; Ellis et al. 2017; Carey 2017a; Geruso et al. 2019; Van Kleef et al. 2019; Shmueli & Nissan-Engelcin 2013; Ellis et al. 2017; Kauer et al. 2020; McGuire & Van Kleef 2018a).

Mitigating risk-selection incentives is important to safeguard efficiency, accessibility and affordability of health insurance. Risk selection can be defined as 'actions by consumers and health plans to exploit unpriced risk heterogeneity and break pooling arrangements' (Newhouse 1996), which can have several unfavorable effects. Firstly, risk selection might reduce efficiency in production if selection is a more effective way of reducing costs than being responsive to the preferences of consumers. Secondly, risk selection might threaten the quality of health care (Glazer & McGuire 2000; Van de Ven et al. 2015). This can happen when health insurers engage in selection via health plan design. If consumers can (to some extent) anticipate which and how much of a service covered by health insurance they will use and when they take this into account when choosing a health plan, they will be sensitive to differences in health plan design regarding those services. Consequently, health insurers can influence consumers' choice for a health plan through the design (Ellis & McGuire 2007; Ellis et al. 2013b; Han & Lavetti 2017; McGuire et al. 2014). By deliberately not contracting/investing in good quality health care for unprofitable (chronically ill) groups, health care providers offering high-quality health care for chronically ill might not get contracted. As a result, quality of care will suffer (Van Kleef et al. 2013a). Thirdly, risk selection might cause price distortions and inefficient sorting

of consumers across health plans, which could threaten both fairness and allocative efficiency of health insurance systems. When high-risk individuals concentrate in other health plans than low-risk individuals, premium variation across plans might not just reflect variation in plan value but also the effect of selection. Such selection-driven premium variation can distort the price/quality tradeoff consumers make when choosing a health plan. Moreover, it threatens the level-playing field for health insurers and reduce the cross-subsidization between high and low risk individuals (Van de Ven & Ellis 2000; Glazer & McGuire, 2000; Van Kleef et al. 2013a; Van Kleef et al. 2013b; Van Kleef et al. 2019; Van de Ven et al. 2015).

Predictable profits and losses after risk equalization imply *incentives* for risk selection for health insurers. While the presence of risk selection itself is very difficult to demonstrate, some studies have found indications of health insurers engaging in risk selection. Ellis et al. (2013b) and Decoralis & Guglielmo (2017) showed that health insurers engage in risk selection to deter unprofitable consumers. Geruso et al. (2016) specifically showed that in the context of the Health Insurance Exchanges in the US, health insurers deliberately distort reimbursement of specific prescription drugs for which they found that the individuals using them are predictably unprofitable. In the same context, Lavetti & Simon (2016) and Han & Lavetti (2017) found evidence of risk selection through reimbursement of prescription drugs, and Shepard (2016) showed that relatively many high-risk individuals left a health plan after this plan dropped a hospital offering high-quality health care from its contracted network. Finally, in the Dutch context health insurers have explicitly stated that they are reluctant to actively invest in health care for specific groups of chronically ill that are undercompensated by the risk equalization model (KPMG 2014; KPMG 2020).

In general, there are several options to mitigate risk-selection incentives in competitive regulated health insurance markets. A first option is to improve the risk-equalization model by adding new or better risk adjusters, and/or by modification of the estimated payment weights. Research has shown that payment weight modification is a particularly promising option that can yield better outcomes in terms of predictable profits and losses (Glazer & McGuire 2002; Van Kleef et al. 2017b). A second option is to use a form of risk sharing. Risk sharing transfers some of the financial responsibility of the health plan to the regulator (usually the government), who retrospectively reimburses health insurers for (some of) the actual spending (Van de Ven & Ellis 2000; McGuire & Van Kleef 2018c). Risk sharing can be implemented in many different forms, of which proportional risk sharing, reinsurance, high-risk pooling, and risk corridors are most common. A drawback of risk sharing is that

while it mitigates risk-selection incentives, it also reduces incentives for health insurers to operate efficiently because health insurers are responsible for a smaller share of total spending. That is, health insurers have fewer incentives to negotiate good prices and/or reduce inefficient spending. The extent to which risk sharing reduces incentives for cost control depends on the specific (design of the) risk sharing scheme and institutional context. Lastly, the regulator might consider allowing some degree of risk rating in which the premium becomes (to some extent) adjusted to the risk of the individual. While this mitigates risk-selection incentives, it can also compromise affordability of health insurance for high-risk individuals.

1.2 A UNIQUE OPPORTUNITY TO EVALUATE AND IMPROVE RISK EQUALIZATION

Evaluating and improving risk equalization requires data that are not explicitly used as a basis for risk adjusters in the risk-equalization model. The reason is that health insurers are completely compensated for the variation in spending between the risk adjusters explicitly included in the model (this is a property of ordinary least squares regression, the common method for estimating risk equalization models). In this light, a unique opportunity has arisen due to the availability of a large and rich health survey (N=384k). This health survey was conducted in 2012 and contains information on physical and mental health as well as lifestyle indicators for the adult population[1]. More specifically, each individual in the survey has self-reported their general health, specific chronic conditions they might have or had been suffering from ever in the past or in the past 12 months and whether they experience any limitations in sight, hearing or physical mobility. They also self-reported specific indicators regarding mental wellbeing like anxiety, depression and loneliness, as well as specific lifestyle indicators such as physical activity, whether and how much they smoke or consume alcohol.

For this thesis, we were able to combine this health survey with the administrative data used to estimate the actual Dutch risk-equalization model. This model is one of the most sophisticated risk-equalization models currently used in the world and includes an extensive set of demographic-, socioeconomic- and morbidity-based risk adjusters. The latter are based on information regarding prior use of prescription drugs, diagnoses from prior hospitalizations, and prior use of medical equipment related to chronic illness. We used the administrative data containing the individual-

1 The health survey data includes adults who were 19 years or older on September 1st, 2012.

level information on spending and risk adjuster variables for every Dutch citizen in 2013 to be able to replicate the actual Dutch risk equalization model of 2016[2].

Health survey data are typically not used as a basis for risk adjusters in risk equalization because these data are not available for the entire population, which is often a requirement (Van de Ven & Ellis 2000; Ellis et al. 2018). While this limits the options to use health survey information directly for improving risk equalization, it still provides ample opportunity to evaluate and improve risk equalization. Van Kleef et al. (2017a) performed an explorative study into this specific large health survey in relation to the Dutch risk-equalization model of 2016. They found that there remain significant predictable profits and losses (i.e., spending variation not compensated for by risk equalization) for selective groups identifiable in the survey. This dissertation builds on the research by Van Kleef et al. (2017a) by further evaluating the risk-equalization model and by studying possible improvements of that model using these health survey data.

1.3 CENTRAL AIM

To guide further improvements of risk-equalization models, the central aim of this dissertation is to study how health survey information can be used to evaluate and improve sophisticated risk equalization. The research question of this dissertation therefore is:

How can health survey information be used to evaluate and improve sophisticated risk equalization in a competitive health insurance market with premium-rate restrictions?

By answering this research question, this dissertation contributes to the research regarding risk equalization and risk selection by specifically studying how health survey information can be used in this context. We do not only use this information to evaluate risk equalization by studying predictable profits and losses for selective groups identifiable in this survey (as was done by Van Kleef et al. 2017a), but especially also to identify source(s) of remaining predictable profits and losses and to study specific strategies to reduce these remaining predictable profits and losses.

2 The risk-equalization model in year t uses spending data from year t-3, which has been made representative for year t.

This dissertation is not the first to examine how health survey information can be used to evaluate risk equalization. Ellis et al. (2013a) have explained individual-level variation in medical spending using a large health survey from Australia and found that health survey information is able to explain some additional individual-level variation in spending compared to the risk equalization model. Also, Lamers (1999) and Stam et al. (2010) have studied the use of health survey information in risk equalization. However, instead of explaining variation in spending, like Ellis et al (2013a), they used a health survey to explain variation in *residual* spending (i.e., actual spending minus the risk equalization payment). Both Lamers (1999) and Stam et al. (2010) also concluded that health survey information has potential to improve risk equalization. This dissertation differs from the above-mentioned studies in several ways. Firstly, the risk equalization model incorporated in this dissertation is more sophisticated, and secondly, compared to Lamers (1999) and Stam et al. (2010) we were able to use a much larger health survey: n=384k versus n=15k in Lamers (1999) and n=23k in Stam et al. (2010). The benefit of such a large health survey compared to the smaller ones is that it substantially reduces the statistical uncertainty surrounding the outcomes of interest. Lastly, we do not only use health survey information to explain residual spending (as the above-mentioned studies did) or to study predictable profits and losses for specific groups (like Van Kleef et al. 2017a), but also to study source(s) of remaining predictable profits and losses and specific strategies to reduce these remaining predictable profits and losses.

This dissertation consists of two parts. The first part examines how health survey information can be used to evaluate risk equalization and identifies driver(s) of predictable profits and losses for selective groups after risk equalization. The second part examines specific strategies using health survey information to reduce the predictable profits and losses identified in the first part.

1.3.1 Evaluating risk equalization using health survey information

As stated above, predictable profits and losses on selective groups caused by premium-rate restrictions lead to risk-selection incentives. Despite sophisticated risk equalization, predictable profits and losses (and thus selection incentives) are not completely eliminated. The first part of this dissertation starts by examining to what extent health survey information can explain variation in residual spending after sophisticated risk equalization. Therefore, the first research question of this dissertation is:

Q1: To what extent does health survey information explain variation in residual spending after sophisticated risk equalization?

We specifically research the risk heterogeneity within the risk adjuster classes of the Dutch risk-equalization model. The rationale for this is that due to the estimation technique (ordinary least squares regression) used for that model, the spending variation between risk adjusters included in the risk equalization model is zero by definition. This implies that the spending variation between risk adjuster classes is compensated for completely and that any spending variation left after risk equalization (i.e., predictable profits and losses) is caused by spending variation *within* risk adjuster classes. To examine the extent to which spending variation within risk adjuster classes is predictable, we use the health survey information. More specifically, we developed a prediction model to explain and predict residual spending using the health survey information. We examined the correlation between the actual and predicted residual spending and the mean absolute value of predicted residual spending. In a situation where risk equalization would completely compensate for the *predictable* spending variation between individuals, both metrics will not yield statistically significant results.

Most studies researching risk-selection incentives typically study predictable profits and losses for one contract period (e.g., 1 year in the Netherlands). However, a multiple contract period perspective might provide additional insights when evaluating selection incentives. The reason for this is that, firstly, due to switching barriers (e.g., switching costs and health insurance illiteracy) consumers do not necessarily switch health insurance at the end of every contract period, leading to suboptimal choices and status quo bias in health insurance (Abaluck & Gruber 2011; Duijmelinck et al. 2015; Handel 2013; Handel & Kolstad 2015; Handel et al. 2018; Heiss et al. 2016; Ketcham et al. 2012; Kling et al. 2012; Laske-Aldershof 2004; Van Vliet 2006). Secondly, when the (un)profitability of a group increases (decreases) over time, this group might be more (less) attractive to a health insurer than suggested by the one-year (un)profitability. The (un)profitability might increase or decrease because predictable profits and losses may or may not persist over time. On the one hand, spending is likely to increase due to deteriorating health. On the other hand, the risk-equalization payment is also likely to increase. Depending on how actual and predicted spending for selective groups develop over time, the (un)profitability of these groups may increase or decrease. The extent to which the (un)profitability of specific groups increase or decrease over time is an empirical question that translates as follows in our second research question:

Q2: To what extent does the (un)profitability of specific subgroups identifiable in the health survey information change over subsequent contract periods?

Predictable profits and losses may or may not increase over time, depending on how spending and predicted spending by the risk-equalization model evolve. To answer this research question, we examined the predictable profits and losses for selective groups identified in the health survey information for multiple contract periods.

In addition to the first two research questions, we examined a specific possible driver of remaining predictable profits and losses for selective subgroups. Obtaining insight into the drivers of predictable profits and losses is important to mitigate risk-selection incentives. One potential driver is end-of-life spending (defined in this thesis as spending in the last 1-5 years of life). Research has shown that spending for individuals who are in the last phase of their life tends to be higher compared to those who are not (Polder et al. 2006; Van Vliet & Lamers, 1998). In addition, a significant share of total health care spending can be attributed to end-of-life spending (Shmuelli et al. 2010; Stooker et al. 2001). End-of-life spending may contribute to remaining predictable profits and losses when individuals near the end of their life are overrepresented in specific subgroups used for evaluating selection incentives and the risk equalization model does not adequately compensate for their high spending.

While the sophisticated Dutch risk-equalization model includes many risk adjusters that compensate for high spending related to age, gender and health, this model does not include a risk adjuster that explicitly flags people who are near the end of their life (Van Kleef et al. 2018). However, the end-of-life stage is likely to correlate with existing morbidity indicators in the risk-equalization model because much ex-post spending on the deceased can be explained by ex-ante spending on the sick (Einav et al. 2018). Therefore, remaining predictable profits and losses for selective groups may or may not be driven by end-of-life spending. The third research question therefore is:

Q3: To what extent are predictable profits and losses after sophisticated risk equalization for selective groups identifiable in the health survey information driven by end-of-life spending?

To answer this research question, we simulated the predictable profits and losses for selective groups identified in the health survey information under sophisticated risk equalization with and without end-of-life spending while correcting for the overall spending difference between these two situations.

1.3.2 Improving risk equalization using health survey information

In contrast to the first three research questions which focus on the use of health survey information for the evaluation of selection incentives, research questions four and five focus on strategies to mitigate risk selection incentives using that same information. While health survey information has been shown to be predictive of residual spending after risk equalization, it is often not suitable for *direct* use in risk equalization in the form of risk adjusters due to feasibility challenges and bias (Ellis et al. 2018). For example, collecting this information for the entire population is often considered too costly and time consuming (Van de Ven & Ellis 2000; Ellis et al. 2018). It can, however, be used in risk equalization *indirectly*. A possible option to do so, is through the method of constrained regression. This is an alternative way of determining the payment weights of the risk-equalization model. Conventionally, risk-equalization models are estimated using ordinary least squares regression (OLS). Given a set of risk adjusters, OLS results in payment weights (i.e., the regression coefficients) that minimize the sum of squared residuals. Constrained regression also results in payment weights that minimize the sum of squared residuals, but conditional on a pre-specified constraint. An example of such a constraint is a pre-specified profit/loss (for instance zero) for a specific subgroup, like those with a fair or (very) poor self-reported general health. Van Kleef et al. (2017b) have researched this specific strategy in the context of the Dutch risk-equalization model and found that constrained regression has the potential to improve the payment fit (i.e., the predicted spending by risk equalization) for some groups but deteriorates the payment fit for other groups. Unlike Van Kleef et al. (2017b), we use health survey information to determine the constraints and study the effect of different constraints for the predictable profits and losses for selective groups identified in the health survey information. The fourth research question therefore is:

Q4: To what extent can constrained regression with constraints based on health survey information improve sophisticated risk equalization?

Another possibility to use health survey information indirectly in risk equalization is to implement high-risk pooling. This is a form of risk sharing between the regulator and the health insurer in which the health survey information is used to assign individuals to the high-risk pool. While the regulator cannot use health survey information in risk equalization because it is not available for the entire population, a health insurer is not restricted by these requirements. Health insurers may use this kind of information for their own risk assessment and, in doing so, find indications of predictable profits and losses, creating risk selection incentives. Through high-

risk pooling a regulator can take advantage of this by enabling health insurers to assign specific individuals/subgroups with high expected losses to a high-risk pool for which they get compensated. This way the predictable losses found for specific high-risk individuals identified through the health survey could be reduced, which mitigates risk selection incentives. The fifth research question therefore is:

Q5: To what extent can high-risk pooling based on health survey information improve sophisticated risk-equalization?

We addressed this question by simulating the assignment of individuals to a high-risk pool using the health survey information from the perspective of the health insurer. We evaluated the predictable profits and losses for selective groups identifiable in the health survey information under five different high-risk pool sizes. In addition, we quantified the tradeoff between incentives for risk selection and incentives for cost control for each of these five pool sizes.

1.4 STRUCTURE OF THIS THESIS

This dissertation is structured as follows. Part one focusses on the use of health survey information in the evaluation of risk equalization. More specifically, chapter two answers Q1 by explaining and predicting residual spending using health survey information. Chapter three answers Q2 by evaluating predictable profits and losses for selective groups identifiable in the health survey information over multiple consecutive contract periods and chapter four answers Q3 by researching to what extent predictable profits and losses after risk equalization are driven by end-of-life spending.

Part two focuses on the use of health survey information for improving risk equalization. Chapter five answers Q4 by examining constrained regression based on health survey information and chapter six addresses Q5 by examining high-risk pooling based on health survey information. The conclusions of chapters two to six are summarized in chapter seven and discussed in chapter eight. In addition, chapter eight highlights important implications for policy as well as important topics for further research.

I

Using health survey information for the evaluation of risk equalization

2

Examining unpriced risk heterogeneity in the Dutch health insurance market

With R.C. van Kleef & F. Eijkenaar

Based on: Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. (2018). Examining unpriced risk heterogeneity in the Dutch health insurance market. European Journal of Health Economics. https://doi.org/10.1007/s10198-018-0979-x

ABSTRACT

A major challenge in regulated health insurance markets is to mitigate risk selection potential. Risk selection can occur in the presence of unpriced risk heterogeneity, which refers to predictable variation in health care spending not reflected in either premiums by insurers, or risk equalization payments. This paper examines unpriced risk heterogeneity within risk groups distinguished by the sophisticated Dutch risk equalization model of 2016. Our strategy is to combine the administrative dataset used for estimation of the risk equalization model (n=16.9m) with information derived from a large health survey (n=387k). The survey information allows for explaining and predicting residual spending of the risk equalization model. Based on the predicted residual spending, two metrics are used to indicate unpriced risk heterogeneity at the individual level and at the level of certain (risk) groups: the correlation coefficient between residual spending and predicted residual spending, and the mean absolute value of predicted residual spending. The analyses yield three main findings: (1) the health survey information is able to explain some residual spending of the risk equalization model, (2) unpriced risk heterogeneity exists both in morbidity and in non-morbidity groups, and (3) unpriced risk heterogeneity increases with predicted spending by the risk equalization model. These findings imply that the sophisticated Dutch risk equalization model does not completely remove unpriced risk heterogeneity. Further improvement of the model should focus on broadening and refining the current set of morbidity-based risk adjusters.

2.1 INTRODUCTION

Many countries have based their health insurance system on principles of regulated competition (Van de Ven et al. 2003). In these systems, health insurers compete on price (i.e. the premium) and quality (e.g. in terms of the contracted provider network) within a regulatory framework set by the government. This regulatory framework aims to achieve public goals, such as individual affordability and accessibility. Common regulatory measures include standardization of the benefits package, premium-rate restrictions, open enrollment and risk equalization (Enthoven & Van de Ven 2007; Van de Ven et al. 2013).

One of the main challenges in regulated health insurance markets is to avoid risk selection (Van de Ven et al. 2003; McGuire et al. 2014; Ellis et al. 2013b; Newhouse et al. 2012). Risk selection has been defined as 'actions by consumers and health plans to exploit *unpriced risk heterogeneity* and break pooling arrangements' (Newhouse 1996, emphasis added). Unpriced risk heterogeneity refers to predictable variation in health care spending not reflected in either premiums by insurers, or in risk equalization payments. The extent to which 'unpriced risk heterogeneity' is present in regulated health insurance markets depends heavily on the specific regulations in place. On the one hand, premium-rate restrictions, standardization of the benefits package and open enrollment introduce or increase unpriced risk heterogeneity. On the other hand, risk equalization reduces unpriced risk heterogeneity by compensating insurers for predictable variation in medical spending (Newhouse et al. 2012; McWilliams et al. 2012; Van de Ven & Ellis 2000, Van Kleef et al. 2019). This paper analyzes unpriced risk heterogeneity in the Dutch health insurance market.

Minimizing unpriced risk heterogeneity is a central objective in regulated health insurance markets, because risk selection has several unfavorable effects. First, risk selection might lead to inefficient health plan design (Glazer & McGuire 2000). For example, insurers do not have incentives to improve (or even maintain) the quality of the contracted care for subgroups that are known to be unprofitable. Secondly, risk selection might lead to price distortions and result in inefficient sorting of consumers across health plans (Einav & Finkelstein 2011). For example, if unprofitable individuals (e.g. those with particular pre-existing conditions) tend to sort into high-quality plans, the incremental premium for these plans does not only reflect the better quality of these plans but also captures some unpriced risk heterogeneity, which distorts consumers' price/quality tradeoff when choosing a health plan. Thirdly, risk selection may reduce efficiency in production if it is a more effective way of reducing costs than negotiating and contracting efficient care. Finally, risk

selection may reduce cross-subsidization from low-risk to high-risk individuals when these risk types are concentrated in different health plans (e.g. high- versus lower-quality plans, see previous example). Incomplete cross-subsidization might lead to compromised accessibility and affordability and violates the level playing field for insurers (Van de Ven & Ellis 2000; Van Kleef et al. 2013a; Van Kleef et al. 2016; Van Kleef et al. 2019).

Although risk equalization systems have become more sophisticated over the past decades, they still do not completely compensate for predictable variation in medical spending. Consequently, given premium rate restrictions, unpriced risk heterogeneity is still present (Van Kleef et al. 2017a; McGuire et al. 2014; Ellis et al. 2013b; Brown et al. 2014; Newhouse et al. 2015; Van de Ven et al. 2003; Kautter et al. 2014). The Dutch risk equalization model, for instance, has been greatly improved over the last decade but still leads to significant under- and overcompensations on specific groups (Van de Ven 2011). For example, in 2008, the subgroup of individuals who reported a fair or (very) poor health status in the prior year (23 percent of Dutch population) was undercompensated by on average 607 euros per person per year. As a result of the introduction of new risk adjusters, however, the average undercompensation for this group reduced to 390 euros in 2016. The same pattern can be observed for other subgroups. So, despite marked improvements of the Dutch model, some unpriced risk heterogeneity remains (Van Kleef et al. 2012; Van Kleef et al. 2014; Van Kleef et al. 2017a; Van Kleef et al. 2018a).

The current over- and undercompensations suggest that the Dutch risk equalization model does not sufficiently identify the risk profile of individuals. More specifically, the morbidity-based risk adjusters in the risk equalization model (see below) might 1) not identify all high-risk individuals and/or 2) not homogenously classify high-risk individuals. Both issues are probably present in the Dutch risk equalization model. Take, for instance, the risk adjuster 'pharmacy-based cost groups' (PCGs), which classifies individuals in morbidity groups based on the use of prescribed drugs (related to chronic illness) in the prior year. For most PCGs individuals must have used at least 181 defined daily doses (DDD) in order to be classified in a relevant PCG (Van Kleef et al. 2018a). On the one hand, PCGs might not identify all people with a particular chronic condition because some of these people might have used less than 181 DDDs in the prior year. On the other hand, PCGs might be heterogeneous in the sense that individuals who use 365 DDDs of a specific drug may be sicker than those who used slightly more than 181 DDDs. Because of both reasons, insurers might not receive the right compensation for specific subgroups.

This paper further examines unpriced risk heterogeneity in the Dutch basic health insurance market. Since premiums for the basic health insurance in the Netherlands are community-rated per health plan, the risk equalization scheme is the main factor influencing unpriced risk heterogeneity. By studying unpriced risk heterogeneity, directions for further improvement of the risk equalization model may emerge, which in turn could further mitigate potential for risk selection. For example, unpriced risk heterogeneity within morbidity groups (as identified by the risk adjusters in the risk equalization model) might call for refinement of existing morbidity-based risk adjusters, while unpriced risk heterogeneity in non-morbidity groups (as identified by the risk adjusters in the risk equalization model) might call for a broader set of morbidity-based risk adjusters.

Identification of unpriced risk heterogeneity requires 'external' information on health risk, i.e. risk indicators that do not explicitly serve as risk adjusters in the risk equalization model. Since the Dutch risk equalization model is estimated by Ordinary Least Squares (OLS) regression, the residual spending for risk classes explicitly included in the model is zero by definition, implying that the variation in spending between these classes will be compensated for completely. Of course, (some of the) variation in spending *within* risk classes will remain. Without external information on risk, however, it is impossible to determine to what extent this variation in spending is predictable.

In this study, the administrative data (2013) that were used to calculate the coefficients of the risk equalization model of 2016 are enriched with external data from a large health survey administered in 2012 (N≈387.000). The administrative data are used to replicate the Dutch risk equalization model of 2016 and to determine individual-level residual spending. Subsequently the health survey data are used to develop a model to explain and predict individual-level residual spending. Unpriced risk heterogeneity is then examined using two metrics: (1) the correlation between (actual) residual spending and predicted residual spending across risk classes and (2) the mean absolute value of predicted residual spending generated by this prediction model. The latter is calculated for the entire sample as well as for specific risk classes distinguished by the risk equalization model.

This study is not the first to use a large health survey to explain individual-level variation in medical spending. Ellis et al. (2013a) use a large health survey from Australia to assess the added value of health survey information in explaining individual-level variation in medical spending. Our study differs from Ellis et al. (2013a) in that it investigates the added value of health survey information in explaining

variation in *residual spending of the risk equalization model*, the difference thus being the incorporation of a risk equalization model. Our approach is similar to that of Lamers (1999) and Stam et al. (2010) who also studied the added value of health survey information in explaining residual spending of the risk equalization model. The risk equalization model used in our study, however, is more advanced and incorporates more information. Another difference with Lamers (1999) and Stam et al. (2010) is that the sample size of our health survey is much larger: 387k versus 15k in Lamers (1999) and 23k in Stam et al. (2010).

The aim of this paper is twofold. The first objective is to study the added value of information derived from a health survey in explaining residual spending of the Dutch risk equalization model 2016. The second objective is to examine unpriced risk heterogeneity within risk classes distinguished by the Dutch risk equalization model.

This paper is organized as follows. In Section 2.2 a brief description of relevant aspects of the Dutch health insurance market is given. In Section 2.3 the data and methods are explained, followed by the results in Section 2.4. Section 2.5 discusses these findings and Section 2.6 concludes.

2.2 THE DUTCH HEALTH INSURANCE MARKET

This section briefly describes the most relevant aspects of the Dutch health insurance system. For a more comprehensive overview, see Van Kleef et al. (2018a) and Enthoven & Van de Ven (2007). Since this paper investigates unpriced risk heterogeneity under the Dutch risk equalization model of 2016, the following description focuses on the situation of 2016.

The analyses in this paper focus on the Dutch basic health insurance. In addition to the basic health insurance, there is a public insurance program for long-term care and a supplementary health insurance for health care services not included in the basic health insurance. The basic health insurance covers, among others, primary care, pharmaceutical care, inpatient and outpatient hospital care, and mental health care. For mental health care a separate risk equalization model is applied which will not be included in the analyses. Instead, this paper focuses on the risk equalization model for curative somatic care, which comprises about 90 percent of total medical spending under the basic health insurance (Van Kleef et al. 2018a).

The government is responsible for the development and improvement of the risk equalization model. In practice, insurers receive a contribution based on the risk characteristics of their insured from a risk equalization fund. In addition to the community-rated premium paid to their insurer, insured pay an income-related contribution to the risk equalization fund, often through their employer (Enthoven & Van de Ven 2007; Van de Ven & Schut 2011).

The risk equalization model predicts medical spending using individual risk characteristics like age and gender, region, socioeconomic status, source of income and health indicators. The latter include seven classifications related to morbidity. The first classification comprises the Pharmacy-based Cost Groups (PCGs), which consists of 33 classes based on people's use of medication in the previous year (see above). A person can be classified in multiple PCGs; individuals who do not reach the predetermined DDD threshold for the relevant pharmaceuticals are categorized in a separate class, i.e. 'No PCG' (Van Kleef et al. 2018a).

A second morbidity classification comprises the Diagnoses-based Cost Groups (DCGs), i.e. 15 classes based on specific inpatient and outpatient hospital diagnoses from the previous year. Insured with multiple diagnoses are categorized in one class only, i.e. the one with the highest residual spending. People without any of the selected diagnoses are categorized in a separate category, i.e. 'No DCG' (Van Kleef et al. 2018a).

A third classification consists of the Multiple-year High Cost Groups (MHCGs) which comprises 7 classes based on the level of spending for curative somatic care in the previous three years. The underlying assumption is that individuals with multiple-year high costs most likely suffer from a chronic illness. Individuals are categorized in one class only, i.e. the class with the highest spending. Individuals that are not classified in one of the 7 MHCG classes are classified in a separate category, i.e. 'No MHCG' (Van Kleef et al. 2018a).

Another classification comprises the Durable Medical Equipment Cost Groups (DMECGs). This risk adjuster classifies individuals on the basis of their use of specific durable medical equipment in the previous year, related to chronic conditions. Individuals are classified in one class only, i.e. the one with the highest residual spending. Again, those without a DMECG are classified in a separate class, i.e. 'No DMECG' (Van Kleef et al. 2018a).

The last three classifications are all based on prior-year spending for specific types of health care, i.e. physiotherapy, geriatric rehabilitation care and home care. The classifications based on physiotherapy and geriatric rehabilitation care spending both include 2 classes: yes/ no spending in the previous year. The classification based on home care spending includes 7 classes; individuals are categorized in one class only, which is the class with the highest spending (Eijkenaar & Van Vliet 2017; Van Kleef et al. 2018a).

2.3 DATA & METHODS

This section describes the data and methods used 1) to study the added value of information derived from a health survey in explaining residual spending of the Dutch risk equalization model 2016 and 2) to examine unpriced risk heterogeneity within risk classes distinguished by the Dutch risk equalization model 2016.

As mentioned in Section 2.1, identification of unpriced risk heterogeneity requires 'external' data (i.e. information about health risk that does not serve as a basis for risk adjuster variables). The reason is twofold. First, the mean residual spending for risk classes in the risk equalization model is zero by definition, a property of OLS. This implies that the variation in spending between these classes will be compensated for completely. Of course, within risk classes variation in spending will remain, but without external information it is impossible to determine to what extent this variation in spending is *predictable*. Second, greater variation in spending within specific risk classes compared to others does not automatically indicate greater unpriced risk heterogeneity in these classes. This study relies on external information from a health survey conducted among a large sample of the adult Dutch population.

Two datasets are used in this study. First, we use administrative data containing individual-level information on medical spending and risk adjusters for all citizens with a basic health insurance in 2013 (n=16.9 million). Second, health survey data from Statistics Netherlands are used containing information on physical and mental health as well as on lifestyle for 387,195 individuals[3]. The health survey data are restricted to individuals of 19 years or older (on September 1, 2012) who do not live in an institution. The sample results from a combination of three surveys held in

3 Public Health Monitor (2012) of the Community Health Services, Statistics Netherlands and the National Institute for Public Health and the Environment.

2012, i.e. the elderly monitor (65 years and above), the adult monitor (19-64) and the health monitor (all ages) (Volksgezondheidenzorg.info). The two datasets are merged at the individual level using the citizen service number (i.e. personal ID number assigned to every Dutch citizen by the government). To protect individuals' privacy, citizen service numbers were anonymized by a trusted third party before the datasets were made available for this research (Van Kleef et al. 2017a). All analyses in this study (except the estimation of the risk equalization model), are conditional on the individuals who participated in the health survey and who successfully merged with the administrative data (see below).

We address the research objectives in four steps, which are explained in more detail below. First, we test the representativeness of the sample. Second, we develop a prediction model to explain residual spending with the information from the health survey. We use this model to make a prediction of residual spending for all individuals included in the health survey. Third, we construct several groups for analyzing within-group risk heterogeneity not explained by the risk equalization model. Finally, we apply specific metrics to examine unpriced risk heterogeneity both at the individual level and at the level of specific groups.

2.3.1 Step 1: Testing the representativeness of the sample

In this first step the sample is compared with the adult population in terms of population frequencies for both risk classes included in the risk equalization model and deciles of spending (Table 2.1 and Figure 2.1). In addition, the samples are compared in terms of mean spending per decile of spending (Figure 2.2).

Of the 387,195 respondents of the health survey, 384,004 respondents successfully merged with the administrative dataset. The match is not 100% due to death and migration of citizens. Unfortunately, many records in the sample contain missing values for one or more crucial items in the health survey. After removing these records, 228,944 records remained for analysis.

Table 2.1 shows the prevalence for several risk classes included in the risk equalization model. The last column presents the prevalence for the total adult population with a health plan in 2013. The adjacent column shows the prevalence for the sample after removal of the missing values. It appears that the sample is overrepresented by the young, the healthy, the higher educated and high-income people.

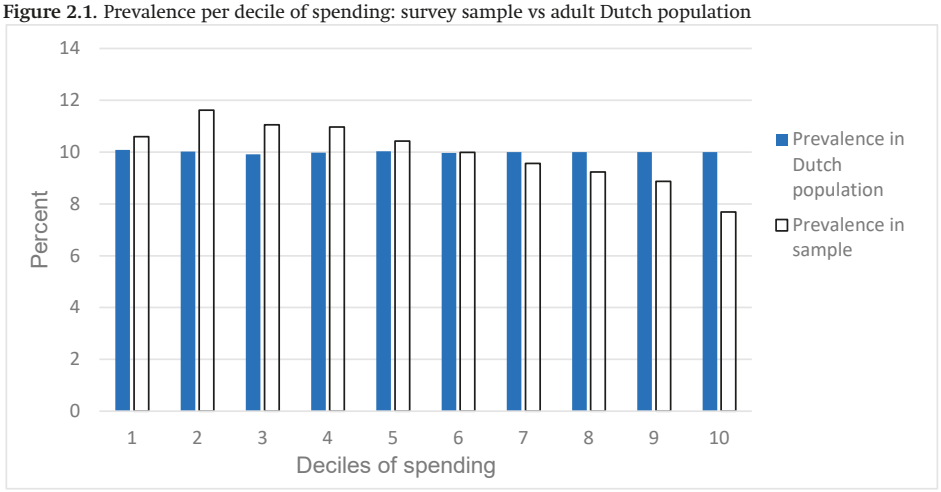
Table 2.1. Prevalence of risk classes: survey sample vs adult Dutch population

		Survey sample	Adult population
Age/Gender	Man, 0-17 year	--	---
	Man, 18-34 year	13.2%*	11.8%
	Man, 35-44 year	9.9%*	8.7%
	Man, 45-54 year	10.6%*	9.8%
	Man, 55-64 year	8.9%*	8.4%
	Man, 65 year and older	8.3%*	10.1%
	Woman, 0-17 year	--	---
	Woman, 18-34 year	13.7%*	11.8%
	Woman, 35-44 year	9.6%*	8.8%
	Woman, 45-54 year	10.3%*	9.8%
	Woman, 55-64 year	8.1%*	8.4%
	Woman, 65 year and older	7.3%*	12.4%
Region	Cluster 1-5	48.5%*	50.2%
	Cluster 6-10	51.5%*	49.8%
Source of income	Younger than 18 or older than 64 year	15.6%*	22.4%
	Full disability benefits	0.3%*	0.4%
	Partial disability benefits	4.1%*	5.4%
	Social security benefits	2.2%*	3.6%
	Student higher education	5.2%*	3.6%
	Self-employed	5.2%*	5.5%
	Higher educated	7.8%*	4.8%
	Other (including salaried employment)	59.5%*	54.4%
Socioeconomic status	Living on an address with >15 people (SES-class 0)	0.1%*	1.1%
	Income decile 1-2 (SES-class 1)	15.5%*	19.7%
	Income decile 3-4 (SES-class 2)	18.5%*	19.8%
	Income decile 5-7 (SES-class 3)	31.9%*	29.8%
	Income decile 8-10 (SES-class 4)	34.1%*	29.6%
Pharmacy-based cost groups	Categorized in at least one PCG	20.1%*	24.1%
Diagnosis-based cost groups	Categorized in a DCG	9.8%*	11.5%
Multiple year high cost groups	Categorized in a MHCG	5.4%*	7.1%
Durable medical equipment cost groups	Categorized in a DMECG	0.8%*	1.1%
Physiotherapy spending in the previous year	Physiotherapy spending in the previous year	2.3%*	2.6%
Home care spending in the previous year	Home care spending in the previous year	1.1%*	2.6%
Geriatric rehabilitation care spending in the previous year	Geriatric rehabilitation care spending in the previous year	0.1%*	0.3%

* = statistically significant different from adjacent figure (P<0.05).

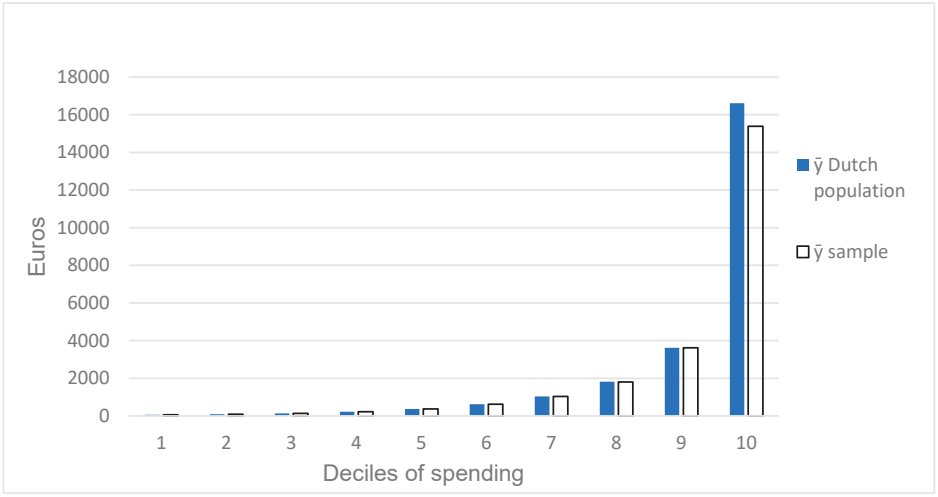
Figure 2.1 compares both groups in terms of the prevalence per decile of spending. The deciles are based on the total Dutch adult population. The bars for the sample show a different pattern than the bars for the population and indicate that the sample is overrepresented by people with low spending. As can be observed in Figure 2.2, however, the average medical spending per decile of spending matches relatively well.

Figure 2.1. Prevalence per decile of spending: survey sample vs adult Dutch population



Note: the deciles on the horizontal axis have been determined on the entire population

Figure 2.2. Average health care spending per decile of spending: survey sample vs adult Dutch population



Note: the deciles on the horizontal axis have been determined on the entire population

2.3.1.1 Recalibrating the survey data

Because the Dutch risk equalization model is estimated by OLS, the mean predicted spending equals the mean spending in the data on which the model is estimated. In other words, the mean residual spending in the population is zero. For the survey sample, however, the mean residual spending equals almost 65 euros per person per year (i.e. an overcompensation). The reasons for this deviation are that 1) the sample is relatively healthy (see Table 2.1 and Figure 2.1) and that 2) apparently, the risk equalization model does not completely correct for this selection bias. In order to correct for this difference in mean residual spending between the sample and the population, we recalibrated the survey data by multiplying the individual-level predicted spending by a factor of 0.967 (i.e. mean spending of 1928 euros in the sample divided by the mean predicted spending of 1992 euros in the sample). After this correction, the mean residual spending in the sample equals zero. Without this correction, our measures of unpriced risk heterogeneity would be affected by the overcompensation on the sample.

2.3.2 Step 2: Building a model to explain and predict residual spending

Next, we built a model to explain and predict residual spending from the risk equalization model. To do so, we first determined the individual-level residual spending e_i by calculating the difference between spending y_i and predicted spending by the risk equalization model $\hat{y}_i$:

$$e_i = y_i - \hat{y}_i \tag{1}$$

Secondly, we use the information from the health survey to explain variance of e_i. This model was developed in three phases: A, B and C. In phase A, we estimated a model including all variables (i.e. 56) available in the health survey as predictors. In order to fully exploit the information from the health survey, in phase B we identified relevant interaction terms. As many interactions are possible, we used a classification tree analysis to identify the statistically significant interaction terms. A classification tree explores higher-order interactions to explain a binary outcome variable (Speybroek 2012). Buchner et al. (2017) have used a similar technique to identify relevant interaction terms for the German risk equalization model. In this study the binary outcome is having a positive (1) or negative prediction error (0) based on the model from phase A. This way the classification tree only yields interaction terms that can explain additional variance in e_i (i.e. variance that is not yet explained by the model from phase A). In phase C the interaction terms were added to the model from phase A and, using stepwise selection, statistically insignificant

variables (*p>0.1)* were dropped. This ultimately led to our final model, which includes 33 variables including three interaction terms (see appendix A). This model is used to predict individual-level residual spending from the risk equalization model:

$$\hat{e}_i = b_0 + b_1 X_{i,1} + \cdots + b_{33} X_{i,33} \tag{2}$$

2.3.3 Step 3: Constructing groups for analyzing within-group unpriced risk heterogeneity

In the third step, two types of risk groups were constructed: 1) groups based on yes/no morbidity, and 2) groups based on deciles of predicted spending by the risk equalization model ($\hat{y}_i$). Morbidity is defined as being classified in at least one of the seven morbidity-based risk adjusters of the risk equalization model (see Section 2.2). Non-morbidity is defined as being classified in none of the seven morbidity characteristics of the Dutch risk equalization model. The deciles of predicted spending result in ten groups based on the predicted spending $\hat{y}_i$. More specifically, we determined deciles of $\hat{y}_i$ for the entire adult Dutch population. These deciles thus order the individuals in the sample according to the predicted spending by the risk equalization model. This is a different way to analyze the relationship between unpriced risk heterogeneity and the risk information included in the risk equalization model.

2.3.4 Step 4: Examining unpriced risk heterogeneity

In the final step, unpriced risk heterogeneity is examined at the individual level and at the level of the groups defined in step 3. Because the sample is overrepresented by healthy individuals, any unpriced risk heterogeneity found probably underestimates the actual unpriced risk heterogeneity. To indicate individual-level unpriced risk heterogeneity, we calculate the R-squared[4] and Cummings Prediction Measure (CPM)[5] of our prediction model from step 2. In addition, we examine the distribution of the (individual-level) predicted residual spending. To indicate unpriced risk heterogeneity per group, two metrics were used. First, the correlation between the individual-level residual spending e_i and individual-level predicted residual spending e_i per group was calculated. This correlation indicates the cohesion between residual spending and predicted residual spending. In a situation where risk equalization completely compensates for predictable variation in spending, residual spending

4 $R^2 = 1 - \frac{\sum_{i=1}^{n}(e_i - \hat{e}_i)^2}{\sum_{i=1}^{n}(e_i - \bar{e})^2}$

5 $CPM = 1 - \frac{\sum_{i=1}^{n}|e_i - \hat{e}_i|}{\sum_{i=1}^{n}|e_i - \bar{e}|}$

will not be predictable. In that case the correlation coefficient will be zero (or at least not statistically significant). Note that because we use a sample of the population, random variation is present. We are, however, interested in systematic variation. This is identified by testing for statistical significance. A statistically significant correlation coefficient indicates unpriced risk heterogeneity.

The second metric is the mean absolute value of predicted residual spending per group j:

$$\overline{|\hat{e}|}_j = \frac{1}{n_j}\sum_{ij}|\hat{e}_i| \tag{3}$$

Again, in a situation where risk equalization completely compensates for predictable variance in spending, residual spending is not predictable. In that case the mean absolute value of predicted residual spending will be (close to) zero. A higher value of $\overline{|\hat{e}|}_j$, indicates more unpriced risk heterogeneity. We deliberately examine the *absolute* values of predicted residual spending and not the *relative* amount of heterogeneity that can be explained by the prediction model (e.g.$\overline{|\hat{e}|}_j$ as a percentage of $\overline{|e|}_j$), as incentives for risk selection are primarily determined by absolute differences between spending and revenues.

2.4 RESULTS

This section presents the results of the empirical analyses. Section 2.4.1 focusses on unpriced risk heterogeneity in the entire sample and Section 2.4.2 on unpriced risk heterogeneity in specific groups. Finally, Section 2.4.3 studies the relationship between unpriced risk heterogeneity and predicted spending by the risk equalization model.

2.4.1 Unpriced risk heterogeneity in the total sample

When explaining variance in medical spending, the prediction model constructed based on health survey variables (see section 2.3.2) yields an R-squared of 10.2 percent. The same model has an R-squared of 0.48 percent when explaining variance in *residual* spending of the risk equalization model of 2016. This indicates that the risk equalization model already performs well in reducing unpriced risk heterogeneity. This conclusion is reinforced when comparing the above mentioned R-squared values to the R-squared of the risk equalization model itself: 28.1 percent. This also shows that the information in the health survey is able to explain only a portion of the variance in residual spending of the Dutch risk equalization model of 2016. The

same conclusion arises when we look at the CPM which equals 0.97 percent for the prediction model explaining variance in residual spending of the risk equalization model compared to 30.5 percent for the risk equalization model itself. Still, the results indicate that after risk equalization some unpriced risk heterogeneity remains.

Figure 2.3 shows the distribution of the predicted residual spending resulting from our prediction model ($\hat{e}_i$) in the total sample. If the risk equalization model would perfectly compensate for predictable variation in medical spending, the prediction model would not be able to predict any residual spending. This would result in a very narrow distribution. The wider the distribution, the better the model predicts residual spending and the more unpriced risk heterogeneity is present. Figure 2.3 indeed indicates presence of unpriced risk heterogeneity in the sample. Figure 2.4 shows the same distribution separately for the morbidity group and the non-morbidity group. The distribution is clearly wider for the morbidity group (panel 2.4a) than for the non-morbidity group (panel 2.4b), indicating a higher level of unpriced risk heterogeneity in the morbidity group than the non-morbidity group. This 'width' of the distribution can be quantified using the mean absolute value of predicted residual spending: $\overline{|\hat{e}_i|}$.

Figure 2.3. Distribution of predicted residual spending for the total sample

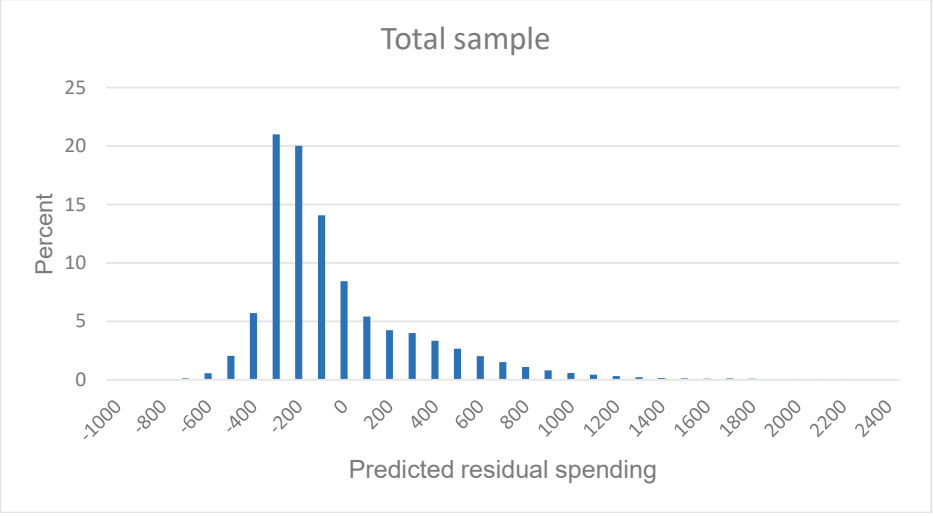
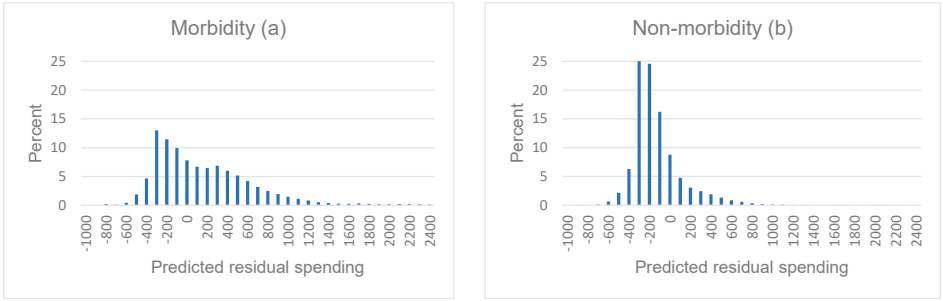


Figure 2.4. Distribution of predicted residual spending for the morbidity group (a) and non-morbidity group (b)



2.4.2 Unpriced risk heterogeneity within risk groups

Table 2.2 provides information on unpriced risk heterogeneity within the morbidity and non-morbidity groups as defined in Section 2.3.3. For all groups we find a statistically significant positive correlation between residual spending and predicted residual spending. Unsurprisingly, the average spending is higher for the morbidity group than for the non-morbidity group. The same is true for the mean absolute value of residual spending $\overline{|e|}_j$, indicating greater unexplained spending variation in the morbidity group as compared to the non-morbidity group (i.e. heteroskedasticity). Note, however, that $\overline{|e|}_j$ not necessarily indicates 'unpriced risk' because variation in residual spending is not necessarily predictable. Therefore, the mean absolute value of *predicted* residual spending $\overline{|\hat{e}|}_j$ is much more interesting. This value indicates the extent to which the health survey variables in the prediction model (which are omitted from the risk equalization model itself, see section 2.3.2) are able to explain residual spending from the risk equalization model of 2016. The last column of Table 2.2 shows that $\overline{|\hat{e}|}_j$ is larger than zero for the morbidity and non-morbidity groups, indicating presence of unpriced risk heterogeneity in both groups. The results show that $\overline{|\hat{e}|}_j$ in the morbidity group is almost twice as high as in the non-morbidity group, implying more unpriced risk heterogeneity in the morbidity group. These findings correspond with Figure 2.4. A deeper look into the constituent elements of the morbidity group (i.e. the PCGs, DCGs, etc.) reveals that the largest values of $\overline{|\hat{e}|}_j$ are found in the classifications based on geriatric rehabilitation care spending and home care spending in the previous year.

Table 2.3 shows the same metrics as Table 2.2, only then for the ten deciles of predicted spending based on the risk equalization model 2016. As described in Section 2.3, these deciles are based on the total adult population, which is why the prevalence in the second column does not equal 10 percent for each class. Kautter et al. (2014) also used deciles of predicted spending in their research, but then to evaluate model performance. In this research the purpose of the deciles of predicted spending is to

Table 2.2. Unpriced risk heterogeneity with regard to morbidity and non-morbidity groups included in the risk equalization model 2016

Group		Estimate size of group	Average spending $\bar{y}_j$	Correlation between e_i and $\hat{e}_i$	$\overline{\lvert e \rvert}_j$	$\overline{\lvert \hat{e} \rvert}_j$
Morbidity	Yes	26.4 %	4675	0.078 ***	3596	383
	No	73.6%	942	0.074 ***	1073	203
PCG	Yes	20.1%	4833	0.085 ***	3519	392
DCG	Yes	9.8%	7008	0.073 ***	5433	490
DMECG	Yes	0.8%	12885	0.120 ***	7790	596
MYHCG	Yes	5.4%	10613	0.101 ***	6744	595
Physiotherapy t-1	Yes	2.3%	7256	0.118***	4764	469
Geriatric rehabilitation t-1	Yes	0.1%	11663	0.085 *	8571	727
Home care t-1	Yes	1.1%	16557	0.135***	9363	765

*= P<0.1, **=P<0.05, ***=P<0.01

order the included groups of the risk equalization model by risk class, enabling us to examine unpriced risk heterogeneity on a different level. For all risk classes a statistically significant correlation between e_i and $\hat{e}_i$ is found. Both the mean absolute value of residual spending $\overline{\lvert \hat{e} \rvert}_j$ and the mean absolute value of predicted residual spending $\overline{\lvert e \rvert}_j$ increases with predicted spending. The latter indicates more unpriced risk heterogeneity among individuals with the highest levels of predicted spending. This corresponds with the results from Table 2.2, which also showed that the groups with high (predicted) spending have higher levels of unpriced risk heterogeneity.

Table 2.3. Unpriced risk heterogeneity with regard to 10 deciles of predicted spending based on the risk equalization model 2016

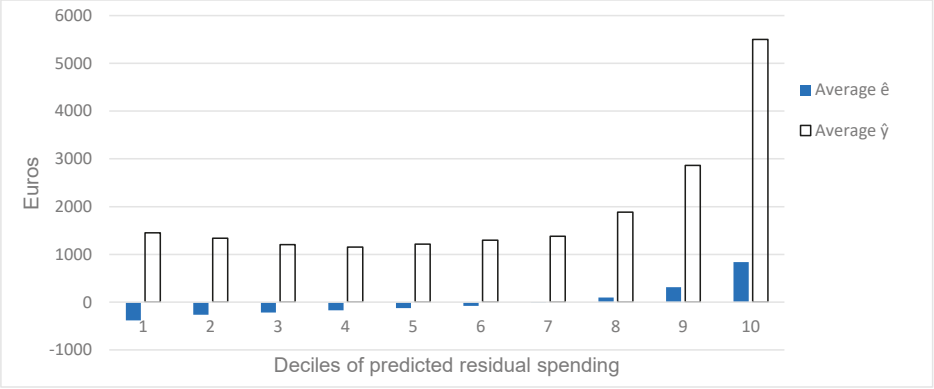
	Size of the group in the sample	Average spending $\bar{y}_j$	Correlation between e_i and $\hat{e}_i$	$\overline{\lvert e \rvert}_j$	$\overline{\lvert \hat{e} \rvert}_j$
Decile 1	14%	422	0.036 ***	467	181
Decile 2	10%	540	0.085 ***	635	189
Decile 3	13%	775	0.073 ***	881	190
Decile 4	9%	984	0.076 ***	1094	198
Decile 5	11%	1085	0.071 ***	1271	210
Decile 6	11%	1448	0.087 ***	1595	217
Decile 7	8%	1679	0.078 ***	1718	262
Decile 8	8%	2290	0.078 ***	2221	291
Decile 9	7%	3590	0.083 ***	3243	362
Decile 10	7%	11183	0.089 ***	7577	623

*= P<0.1, **=P<0.05, ***=P<0.01

2.4.3 Predicted residual spending and predicted spending

Figure 2.5 shows the relation between unpriced risk heterogeneity and predicted spending from a different angle. On the horizontal axis the deciles of predicted residual spending are depicted. Both the mean predicted residual spending $\bar{\hat{e}}_j$ (shaded bars) and the mean predicted spending $\bar{\hat{y}}_j$ (empty bars) per decile j of the risk equalization model are shown. This shows an interesting pattern: the two highest deciles of predicted residual spending also have the highest predicted spending by the risk equalization model. This suggests that for people with the highest predicted residual spending, the risk equalization model already predicts high medical spending, only not high enough.

Figure 2.5. $\bar{\hat{e}}_j$ (mean predicted residual spending) and $\bar{\hat{y}}_j$ (mean predicted spending) per decile j of predicted residual spending



2.5 DISCUSSION

A major challenge in regulated health insurance markets is to avoid risk selection, which can occur in the presence of unpriced risk heterogeneity. Risk equalization models aim to mitigate unpriced risk heterogeneity by compensating insurers for predictable variation in medical spending. In this paper we examined unpriced risk heterogeneity in the Dutch basic health insurance market. Our findings comprise three main conclusions, which are discussed below.

2.5.1 Health survey information indicates unpriced risk heterogeneity

We examined unpriced risk heterogeneity within risk groups included in the risk equalization model. To that end, we merged administrative data on spending and risk characteristics of 2013 for the entire adult Dutch population with rich health survey

data from 2012. With the information from the health survey a prediction model was constructed to explain and predict residual spending from the risk equalization model. The health survey information is able to explain approximately 10 percent of variation in individual-level medical spending. Ellis et al. (2013a) found a similar R-squared of 10 percent for a model based on external health survey information to explain variance in individual-level medical spending in Australia. When explaining variation in *residual* spending (i.e. after application of the Dutch risk equalization model 2016), the R-squared drops to 0.48 percent, given this data, indicating that the health survey information is able to explain a small but non-negligible share of this variation. This confirms that although the Dutch risk equalization model 2016 performs quite well, some unpriced risk heterogeneity - and thus potential for risk selection – remains. These findings are in line with previous research. Recent studies on risk heterogeneity by Newhouse et al. (2015) and Van de Ven et al. (2017) also show there is still selection potential in markets with sophisticated risk equalization models. In addition, Stam et al. (2010), who analyzed the predictive power of self-reported health measures for a risk equalization model that already included several morbidity characteristics (i.e. PCGs and DCGs) also found that these self-reported health measures have added value in explaining medical spending. The R-squared of 0.48 percent found in this study is lower than the incremental change in R-squared of 2 percent, found by Stam et al. (2010). This difference can be explained by improvements of the Dutch risk equalization model over the past decade (Van Kleef et al. 2017a; Van Kleef et al. 2012; Van Kleef et al. 2014).

2.5.2 Unpriced risk heterogeneity is present in both morbidity and non-morbidity groups

Our findings indicate that unpriced risk heterogeneity is present in both the morbidity group and the non-morbidity group included in the risk equalization model. In addition, our findings suggest more unpriced risk heterogeneity in the morbidity group than the non-morbidity group. In line with the latter finding, unpriced risk heterogeneity increases with predicted spending by the risk equalization model. These results lead us to the conclusion that per-person predictable profits and losses (i.e. over- and undercompensations) are larger in the morbidity group than in the non-morbidity group and increase with predicted spending. Apparently, the high-risk group identified by the morbidity indicators in the risk equalization model are to some extent heterogeneous. This calls for further refinement of these indicators, to the extent that remaining unpriced risk heterogeneity in these groups is considered a problem.

When it comes to selection potential, however, not just the per-person predictable profits and losses matter, but also the size of the relevant group. Although the per-person predictable profits and losses are smaller in the non-morbidity group compared to the morbidity group, the former group is larger. Therefore, unpriced risk heterogeneity in this group should not be neglected. It appears that the morbidity indicators do not identify all high-risk individuals and leave some of these people in the non-morbidity group, which calls for extending the set of morbidity indicators.

2.5.3 Relationship between predicted residual spending and predicted spending

Our findings also suggest a relationship between predicted residual spending (following from our prediction model) and predicted spending by the risk equalization model: those with the highest predicted residual spending also have high predicted spending. This shows that the risk equalization model predicts high spending for these individuals, only not high enough. One option to reduce undercompensation for specific groups is to extend the risk equalization model with new risk adjusters that identify these groups (Van de Ven & Ellis 2000). When new or better risk adjusters are not available (in the short run), another option to reduce undercompensation for specific groups is overpaying individuals on the basis of their predicted spending from the risk equalization model. Such overpayment can be realized by, for instance, the use of constrained least squares regression (Glazer & McGuire 2000; Van Kleef et al. 2017b).

2.5.4 Limitations

The findings in this paper must be viewed in the light of some limitations. First, missing values in the health survey data necessitated the exclusion of approximately 150,000 individuals from our analyses. Nonetheless, a substantial sample size of over 200,000 individuals remained. A second limitation may be overfitting as a result of the use of a classification tree to identify relevant interactions between health survey variables (to enrich our prediction model). However, our main aim in identifying interactions was to optimally use the health survey information in explaining residual spending (and thus to indicate remaining unpriced risk heterogeneity), and not to develop potential new risk adjusters to include into the risk equalization model. Moreover, adding the identified interactions to the prediction model only marginally increases the R-squared of this model (i.e. from 0.45 percent to 0.48 percent), suggesting that the non-interaction variables explain the vast majority of the variation in residual spending explained by the prediction model. Thirdly, in this research the information from the health survey was used to explain variation in residual spending of the risk equalization model. Alternatively, the

information from the health survey could have been added to the risk equalization model and this extended model could then have been compared to the original risk equalization model, similar to the approach of Stam et al. (2010). The reason for choosing the first approach lies in the risk of overfitting. Despite having a large sample, the number of people that would be classified in specific categories of the risk equalization model would be too small to yield trustworthy estimates.

Finally, in this research the potential for risk selection has been explored by examining the existence of unpriced risk heterogeneity. It is important to note that the potential for risk selection does not depend on the existence of unpriced risk heterogeneity alone. As recognized in the definition of risk selection by Newhouse (1996), unpriced risk heterogeneity is just one of the conditions that need to be present in order for risk selection to take place. The other conditions relate to 'actions by consumers or health plans' (Newhouse 1996). This includes all possible actions by insurers, regardless of intentions, to exploit unpriced risk heterogeneity as well as the response of consumers to these actions.

2.6 CONCLUSION

This study examined the unpriced risk heterogeneity in the Dutch health insurance market. The analyses yield three main findings: (1) the health survey information is able to explain some residual spending of the risk equalization model, (2) unpriced risk heterogeneity exists in both morbidity and non-morbidity groups, and (3) unpriced risk heterogeneity increases with predicted spending by the risk equalization model. These findings imply that – despite its broad set of morbidity-based risk adjusters – the Dutch risk equalization model 2016 does not completely remove unpriced risk heterogeneity. Further improvement of the model should focus on broadening the current set of morbidity based risk adjusters (to reduce unpriced risk heterogeneity in the large non-morbidity group) and on refinement of the current morbidity-based risk adjusters (to reduce unpriced risk heterogeneity in the morbidity group), if improvement for these groups is desired through risk equalization.

APPENDIX A: Prediction model

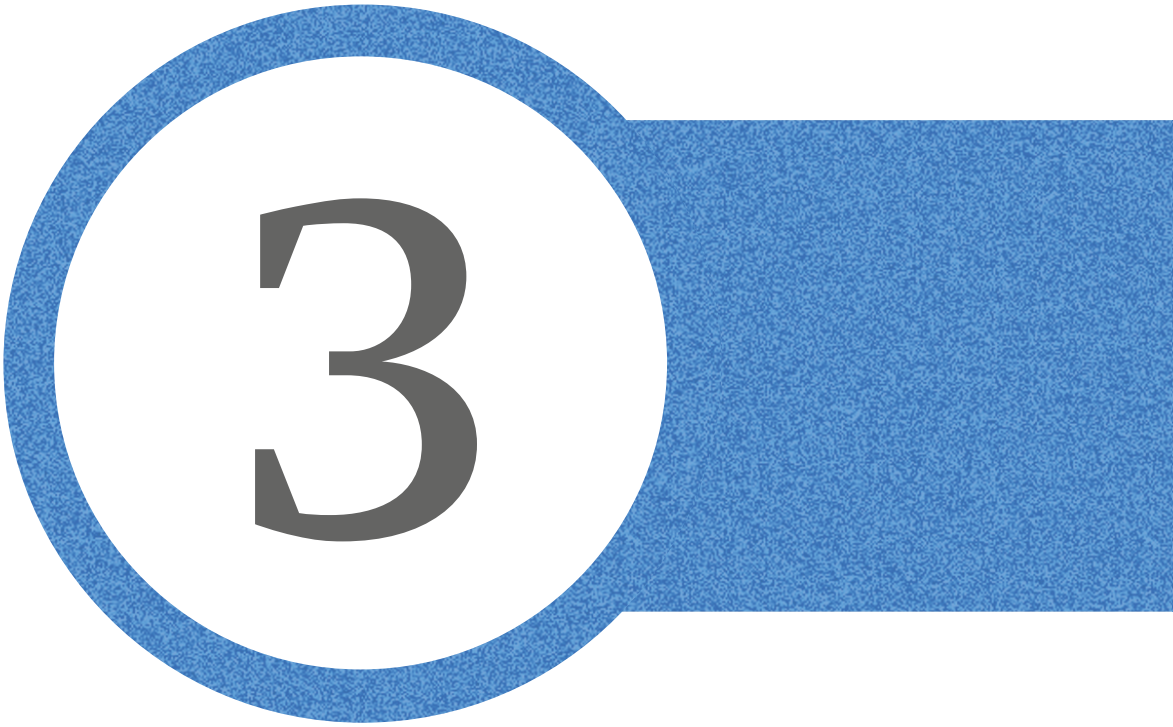
Table A.2.1. prediction model

Variable	Coefficient
Intercept	-29
Fair or (very) poor general health	453
Cancer in the last 12 months	381
Heart condition in the last 12 months	576
Stroke in the last 12 months	1041
Interaction term 1*	2439
Interaction term 2*	1479
Severe/ recurrent disease of intestines in the last 12 months	269
Immigrant of the first generation	-224
Other chronic illness in the last 12 months	153
Sufficient physical activity according to 'fit' norm	-116
3 self-reported conditions	220
Moderate smoker	136
Stroke ever	273
Semi-sufficient physical activity according to 'fit' norm	-82
Loneliness on a social level	-129
OECD limitations in hearing	210
High risk of incurring anxiety disorder or depression	-294
Peripheral artery disease in the last 12 months	232
Heavy smoker	154
Psoriasis in the last 12 months	158
Loneliness, moderately	77
OECD limitations in mobility	121
Interaction term 3*	359
Diabetic	-150
Sufficient physical activity according to 'beweeg' norm	-138
Heavy drinker	92
Severe/recurring dizziness in the last 12 months	139
Semi-sufficient physical activity according to 'beweeg' norm	-100
2 self-reported conditions	67
Severe/recurring condition of back in the last 12 months	-80
High educated	48
Obese	65
OECD limitations in sight	103

All variables are statistical significant at $p<0,1$
R^2 of model is 0.0048
R^2 of model without interactions is 0.0045

*interaction terms:

	Interaction term	N (weighted)
1	chronic illness * OECD limitations in mobility * no asthma 12 months * no low risk of incurring anxiety disorder or depression * no joint inflammation 12 months * cancer 12 months	9966
2	Chronic illness * no OECD limitations in mobility * no severe/recurring condition of neck in the last 12 months * no low risk incurring anxiety disorder or depression* no joint inflammation 12 months * other chronic illness 12 months * cancer 12 months	26416
3	Chronic illness * no OECD limitations in mobility * no severe/recurring condition of neck in the last 12 months * high risk of incurring anxiety disorder or depression* no fair or (very) poor general health	57790


3

Predictable profits and losses in a health insurance market with risk equalization: A multiple-contract period perspective

With R.C. van Kleef & F. Eijkenaar

Based on: Withagen-Koster, A.A., van Kleef, R.C., Eijkenaar, F. Predictable profits and losses in a health insurance market with risk equalization: a multiple-contract period perspective. Submitted

ABSTRACT

Many social health insurance systems rely on 'regulated competition' among insurers to improve efficiency. In the presence of community-rated premiums, risk equalization is an important regulatory feature to mitigate risk-selection incentives in such systems. Empirical studies evaluating selection incentives have typically quantified group-level (un)profitability for one contract period. However, due to switching barriers, a multiple contract period perspective might be more relevant. In this paper, using data from a large health survey (N≈380k) we identify subgroups of chronically ill and healthy individuals in year t and follow these groups over three consecutive years. Using administrative data covering the entire Dutch population (N≈17m), we then simulate the mean per person predictable profits and losses (i.e. spending predicted by a sophisticated risk-equalization model minus actual spending) of these groups over the three follow-up years. We find that most of the groups of chronically ill are persistently unprofitable on average, while the healthy group is persistently profitable. This implies that selection incentives might be stronger than initially thought, underscoring the necessity of eliminating predictable profits and losses for the adequate functioning of competitive social health insurance markets.

3.1 INTRODUCTION

Like in many other countries in the world, the healthcare system in the Netherlands relies on 'regulated competition' among health insurers to simultaneously achieve fairness and efficiency in healthcare financing. In this system, health insurers compete on price and quality within regulatory boundaries set by the government. Crucial regulatory features to help ensure individual accessibility and affordability of health insurance include a standardized benefits package, open enrollment, an insurance mandate, community-rated premiums and a system of risk equalization among insurers (Van de Ven & Ellis, 2000; Van de Ven & Schut 2008;Van de Ven & Van Kleef 2016). The latter, risk equalization, aims to compensate for predictable spending variation in order to mitigate incentives for insurers to engage in risk selection (Van de Ven & Ellis 2000). This is important as risk selection might violate both fairness and efficiency (Glazer & McGuire 2000; Van Kleef et al. 2019; Van de Ven et al. 2015). Even though the Dutch risk equalization model is one of the most sophisticated models currently used in the world, risk-selection incentives towards specific subgroups remain (Van Kleef et al. 2017a; Van Kleef et al. 2019).

The problem of imperfect risk equalization leading to risk-selection incentives is not unique to the Dutch context. Much empirical research has been published on selection incentives under risk equalization models in various other contexts. Typically, these studies analyze predictable profits and losses for certain subgroups for *one* contract period (Glazer & McGuire 2000; Van Kleef et al. 2019; Bauhoff 2012; Buchner et al. 2013; Carey 2017a; Ellis et al. 2013b; is et al. 2017; Geruso et al. 2019; McGuire et al. 2013; McGuire et al. 2014; Shmueli & Nissan-Engelcin 2013; Withagen-Koster et al. 2018). However, a multiple contract period perspective might be more relevant in this regard as consumers do not necessarily switch health plans at the end of every contract period due to barriers like switching costs, health insurance illiteracy and other frictions leading to sub-optimal choices and status-quo bias in health insurance (Abaluck & Gruber 2011; Duijmelinck et al. 2015; Handel 2013; Handel & Kolstad 2015; Handel et al. 2018; Heiss et al. 2016; Ketcham et al. 2012; Kling et al. 2012; Laske-Aldershof et al. 2004; Van Vliet 2006).

Over time, predictable profits and losses may or may not persist. On the one hand, spending is likely to increase over time due to ageing and deterioration of health. On the other hand, risk-equalization payments (i.e. spending predicted by the risk equalization model) is also likely to increase. Especially groups of chronically ill may be flagged by more morbidity indicators used in the risk equalization model and/or morbidity indicators with higher payment weights, but also healthy subgroups

could come with higher risk-equalization payments over time. Thus, depending on the net effect of the development in spending and predicted spending/payments and assuming community-rated premiums, (un)profitability – i.e. predicted spending minus actual spending – of subgroups may increase or decrease over time with a concomitant change in selection incentives as result. If for group g predicted spending increases more than actual spending, (un)profitability of g will increase (decrease) over time implying that g might be more attractive than suggested by one-year (un)profitability, and vice versa. The extent to which (un)profitability of a subgroup increases (decreases) over time is an empirical question.

In this paper, we aim to answer this empirical question by examining the (un) profitability for specific subgroups over multiple contract periods. More specifically, by merging health survey data (n≈380k) on self-reported medical conditions with administrative data (n≈17mln) on both spending and predicted spending derived from a sophisticated risk-equalization model, we simulate group-level (un)profitability over three consecutive contract periods (in our case: years). In our empirical simulation we track patterns in spending and predicted spending over time and calculate the net effect on (un)profitability.

This paper is not the first to study (un)profitability over multiple contract periods in the context of competitive health insurance markets with premium regulation and risk equalization. Van Veen et al. (2016) studied residual spending under the Dutch risk equalization model of 2013 for three consecutive years. They found that there are individuals with a persistent loss and that most of these individuals have multiple long-term diseases. In another study, Kauer et al. (2020) examined residual spending under the Swiss risk equalization model by investigating how many of the individuals with positive (negative) residual spending in the first year also have positive (negative) residual spending in the second year. The authors indeed found that individual-level residual spending is to some extent persistent. Finally, Farid & McGuire (2018) studied extreme under- or overcompensations by risk equalization payments in the Marketplaces in the United States. They too found persistent patterns, especially in the tails of the residual-spending distribution. The important difference between our study and the three abovementioned studies, however, is that we evaluate (un)profitability net of risk equalization over multiple contract periods at the level of specific *subgroups* instead of the individual level. With this group-level approach we fill an important gap in the literature given that many selection actions that are possible in regulated health insurance markets take place at the level of groups rather than individuals (Van Veen et al. 2015b). Examples include group

advertisement, collective arrangements and choices regarding provider network and coverage of out-of-network spending on services related to specific diseases.

This paper is structured as follows. Section 3.2 describes the data and methods. The results are presented in Section 3.3 and Section 3.4 concludes and discusses the findings.

3.2 DATA & METHODS

The aim of this study is to examine the (un)profitability of specific subgroups over multiple contract periods under sophisticated risk equalization. The study is conducted from the perspective of the health insurer in contract period t and examines (un)profitability in year t+1, t+2 and t+3 under the assumption that consumers tend to stay enrolled with the same insurer for multiple contract periods. This section first describes the data, followed by a description of the methods.

3.2.1 Data

We used two primary data sources. First, we used administrative data from the Netherlands for the years 2013, 2014 and 2015. These data come from various administrative sources and contain information on individual-level medical spending and risk adjusters for all Dutch citizens with a basic health insurance in 2013, 2014 and/or 2015 (n≈16.9 million per year). These data were those actually used to estimate the Dutch risk equalization model for somatic care for the years 2016, 2017 and 2018, respectively. The Dutch risk-equalization system comprises separate models for somatic care, mental health care and copayments due to the mandatory deductible. This research focusses exclusively on the somatic model, which concerns about 90 percent of total health care spending covered by the basic benefits package (Eijkenaar et al. 2015; Van Kleef et al. 2018a). The somatic model of 2016 contained the following socio-demographic and morbidity-based risk adjusters: age interacted with gender, region, socioeconomic status and source of income (both interacted with age), Pharmacy-based Cost Groups (PCGs), Diagnosis-based Cost Groups (DCGs), Multiple year High Cost Groups (MHCGs), Durable Medical Equipment Cost Groups (DMECGs), physiotherapy-spending in the prior year, geriatric rehabilitation care spending in the prior year and home care spending in the prior year (see Van Kleef et al. (2018a) for details). In the period 2016- 2018 the model has been updated and slightly altered (See Appendix A).

The second dataset is derived from a health survey conducted in 2012 and contains information on self-reported chronic conditions (see Section 3.3.2) for a sample of 387,195 individuals who were 19 years or older on September 1, 2012[6]. These data come from Statistics Netherlands (Volksgezondheidenzorg.info). Prior research has shown that the survey sample is fairly representative for the adult Dutch population with a basic health insurance in 2013 (Van Kleef et al. 2017a). To further improve the representativeness, we rebalanced the sample by means of a raking procedure that was originally developed by Deming (1943). With this procedure individual-level weights are generated that equalize the frequencies of key variables in the sample to those in the population (Battaglia et al. 2009; Izrael et al. 2000). We included all risk-adjuster classes of the risk equalization model 2016 (see Van Kleef et al. (2018a) for a complete list) as well as a proxy for whether someone had died in 2013 and 18 quantiles of mean total somatic spending into our raking procedure. See Withagen-Koster et al. (2020) for a more detailed description of the rebalancing procedure as well as results on the representativeness of the survey sample before and after rebalancing. After rebalancing, for these variables the frequencies in the health survey sample exactly match those in the population of 2013 (Withagen-Koster et al. 2020).

3.2.2 Methods

We used the health-survey sample to identify groups with an overrepresentation of people in relatively poor or good health in 2012 (i.e. year t), which we then followed over three consecutive contract periods (i.e. t+1, t+2 and t+3). Because we follow groups determined in year t over time, no new cases are added to the identified groups. We then combined the survey data with the administrative data to calculate (un)profitability for the identified groups for each of these years. Specifically, we use the administrative data of 2013, 2014 and 2015 to simulate the (un)profitability in t+1, t+2 and t+3 for the groups identified in data of 2012 (i.e. year t). The outcomes in year t+1 in our study can be used to evaluate the (un)profitability for one contract period, i.e. the conventional procedure.

Our simulation consisted of four steps. First, we defined a sophisticated risk-equalization model. This model comes very close to the actual morbidity-based risk equalization model used in the Netherlands (see Section 3.2.1). To ensure that subgroup-level results are not affected by changes in the model over time, we kept the definitions of the risk-adjuster variables as constant as possible (see Appendix A for a detailed overview of the risk adjusters used in our simulation of predicted

6 Public Health Monitor (2012) of the Community Health Services, Statistics Netherlands and the National Institute for Public Health and the Environment.

spending). Although several small definition differences remain, these are very unlikely to have markedly affected our findings.

In step two, we first recalibrated mean spending in the population of 2014 (t+2) and 2015 (t+3) to the level of 2013 (t+1) to correct for differences in underlying definitions of spending across years. Specifically, we multiplied the individual-level spending for 2014 by 0.95681 (i.e. mean spending of 2013 = 2131 euros divided by mean spending of 2014 = 2227 euros) and for 2015 by 0.93271 (i.e. mean spending of 2013 = 2131 euros, divided by mean spending of 2015 = 2285 euros). Next, we estimated the model defined in step 1 by ordinary least squares (OLS) on total somatic spending in t+1, t+2 and t+3 (i.e. three models in total) to simulate individual-level (un) profitability. Individual-level (un)profitability is defined as spending predicted for individual i by the risk-equalization model minus actual spending for i as observed in our data.

Third, we merged the individual-level results on spending, predicted spending, and (un)profitability with the health survey data using a unique and anonymized (by a trusted third party) individual-level identification key and determined our groups of interest. The health survey data allow for identification of 19 chronic illnesses that respondents could report to have suffered from in the last 12 months (see Appendix C) and 4 chronic illnesses that respondents could indicate to have ever suffered from. Based on these 23 groups we also constructed two additional groups indicating 'no chronic condition' and 'at least one chronic condition'

Finally, we calculated the average (un)profitability for each of the 25 groups identified in step 3 under each of the three models/years, enabling us to investigate group-level (un)profitability over time. For ease of presentation we first recalibrated the spending in the health survey to the administrative data of 2013 (t+1). Because the Dutch risk equalization model is estimated by OLS, the grand mean predicted spending equals the grand mean spending in that dataset, implying a grand mean residual of zero euros. However, for subsamples drawn from that dataset, this is not necessarily the case. The rebalancing procedure (see section 3.2.1) already reduced the mean residual spending in the health survey from -31 euros to -6 euros. To correct for this remaining difference, we recalibrated spending in the sample by multiplying individual-level spending by a factor of 1.002457 (i.e. mean predicted spending of 2460 euros divided by mean spending of 2454 euros) leaving a mean residual of zero euros. This correction simplifies interpretation of our findings: for any mutually exclusive set of subgroups identified in the health survey data the grand mean residual equals zero.

3.3 RESULTS

This section presents the results of our simulations. Section 3.3.1 summarizes some descriptive statistics of the health survey sample and the population in the administrative data. Results regarding group-level spending and predicted spending are presented in Section 3.3.2 and Section 3.3.3 presents the results regarding group-level (un)profitability.

3.3.1 Descriptive statistics

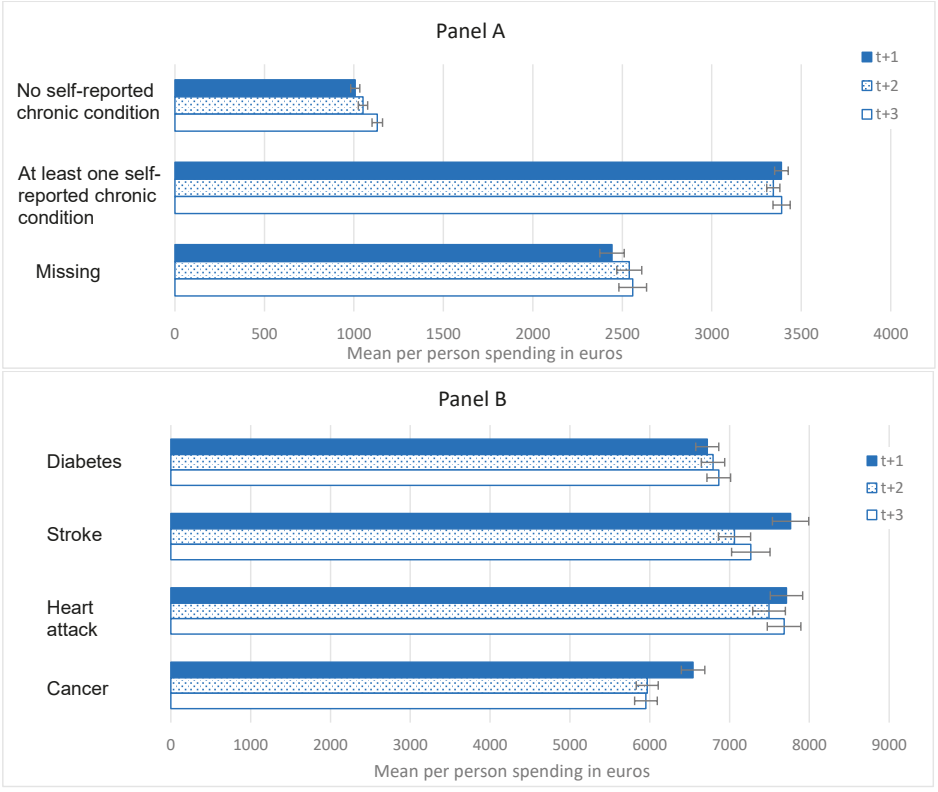
This section summarizes some descriptive statistics of the health survey sample and the population in the administrative data. As shown in Appendix B, the population frequencies of selected risk-adjuster variables in the administrative data match relatively well across the three years, with the small differences in frequencies unlikely to have a notable impact on our results. In addition, the frequencies of the studied subgroups show a small decrease over time. This decrease can be caused by death, migration or unsuccessful merging of the health survey data with the relevant year of the administrative data. For more detailed descriptive statistics and the corresponding figures, see Appendix B.

3.3.2 Group-level spending and predicted spending over multiple contract periods

This section presents the mean (predicted) spending for the identified groups in t+1, t+2 and t+3 and changes therein over time. Note that mean spending between the years has been recalibrated on the population-level. Individual-level fit of the estimated risk-equalization model (used for generating spending predictions for each individual in each year) is roughly similar in the three years: the R-squared changes from 27.9% in t+1 to 28.2% and 27.5% in t+2 and t+3, respectively; Cummings Prediction Measure (CPM) changes from 29.9%, to 30.6% and 30.1% in t+1, t+2 and t+3, respectively.

Figure 3.1 shows the mean actual spending in euros for selected groups. As can be seen in panel a, spending remains relatively constant over time for the group reporting at least one chronic condition, while it increases slightly for the group ‘no chronic condition’. Panel b, which shows the development in spending for specific conditions, shows a relatively constant pattern for diabetes and heart attack, while for stroke and cancer a statistically significant 9% reduction in spending is observed from year t+1 to year t+2. A possible explanation for these differences is the variation in spending over time. For stroke and cancer, it is likely that most of the spending is generated in the first years after diagnosis, while for e.g. diabetes spending can be expected to be much more stable.

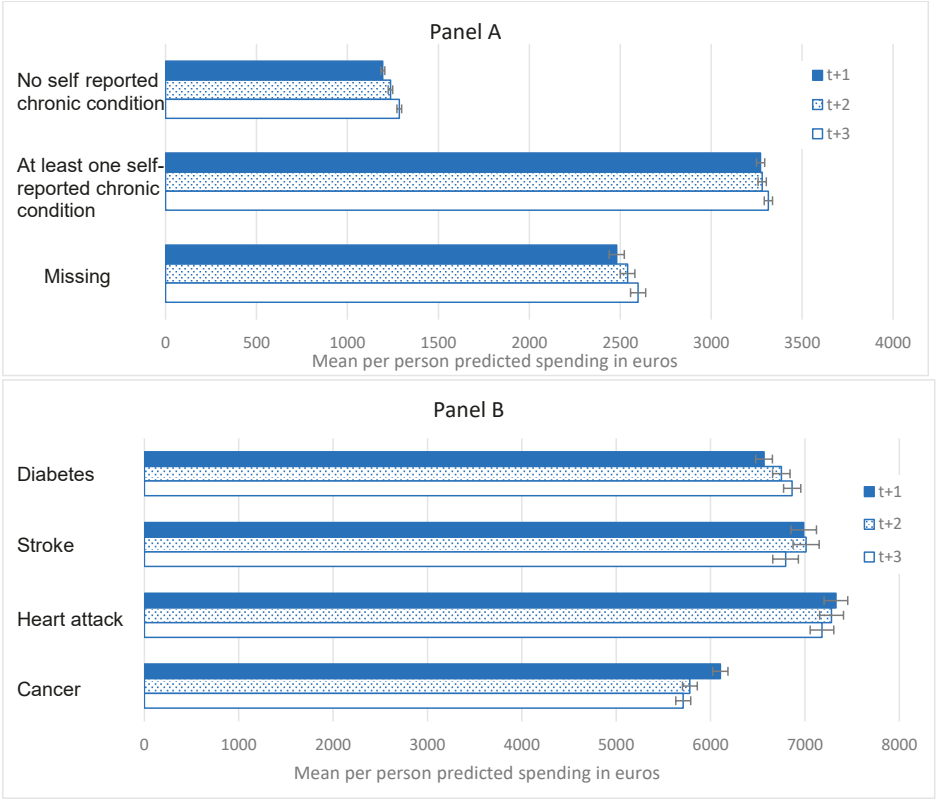
Figure 3.1. Mean spending in euros in years t+1, t+2 and t+3 for individuals who in year t reported to have suffered from no or at least one chronic condition and the group for whom this information is missing (panel a) and for individuals who in year t reported to have ever suffered from diabetes, a stroke, a heart attack or cancer (panel b)



Note: Mean spending for the population in t+2 and t+3 has been recalibrated to the level of t+1. The 95% confidence intervals are calculated as $\overline{y_g} \pm 1.96 * \frac{\sigma_g}{\sqrt{n_g}}$

Figure 3.2 shows the mean spending as predicted by the risk equalization model for the same groups for the three years. (Appendix C shows the results on mean (predicted) spending for the 19 specific chronic conditions.) Panel a shows that the pattern for predicted spending matches the pattern of spending as shown in panel a of Figure 3.1. This is also true for panel b, although for stroke and heart attack the changes are somewhat different (e.g. for stroke predicted spending reduces from t+2 to t+3 whereas actual spending reduces from t+1 to t+2). In addition, although for diabetes predicted spending follows the same (increasing) pattern as compared to actual spending, the increase is larger for predicted spending. For this group, however, the changes from year-to-year are not statistically significant for both spending and predicted spending.

Figure 3.2. Mean predicted spending in euros in years t+1, t+2 and t+3 for individuals who in year t reported to have suffered from no or at least one chronic condition and the group for whom this information is missing (panel a) and for individuals who in year t reported to have ever suffered from diabetes, a stroke, a heart attack or cancer (panel b)



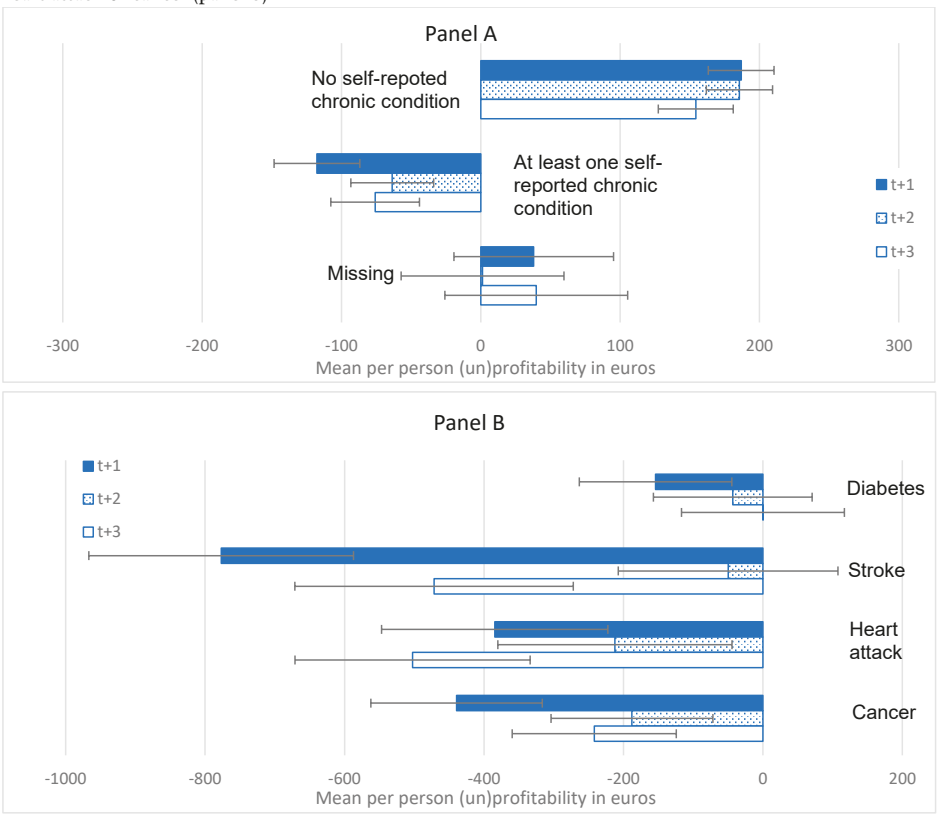
Note: The 95% confidence intervals are calculated as $\overline{y_g} \pm 1.96 * \frac{\sigma_g}{\sqrt{n_g}}$

3.3.3 Group-level (un)profitability over multiple contract periods

This section contains the results of the simulations of (un)profitability in year t+1, t+2 and t+3 after risk equalization for the 25 subgroups identified in the survey sample. (Un)profitability is defined as the difference between the mean predicted spending and the mean actual spending for a subgroup (see Section 3.3.2). Figure 3.3 presents these results for the same groups as shown in Figures 3.1 and 3.2. Panel a shows that the group with no chronic condition is persistently profitable, while the opposite holds for the group with at least one chronic condition. For panel b, the results show that on average, each of the groups diabetes, stroke, heart attack and cancer are unprofitable in year t+1. Except for diabetes, the loss is persistent over time for the other three groups; for these groups, the loss first drops in t+2 before increasing again in t+3. For heart attack, the loss in t+3 even exceeds that of t+1.

For all groups in Figure 3.3 except stroke and cancer, however, the 95% confidence intervals around the average losses overlap, so changes from one year to another could reflect random variation. Except for diabetes, the average loss is statistically significant different from zero in at least year t+1 and t+3, indicating persistence in unprofitability for these groups of chronically ill people.

Figure 3.3. (Un)profitability in euros after risk equalization in years t+1, t+2 and t+3 for individuals who in year t reported to have suffered from no or at least one chronic condition and the group for whom this information is missing (panel a) and for individuals who in year t reported to have ever suffered from diabetes, a stroke, a heart attack or cancer (panel b)

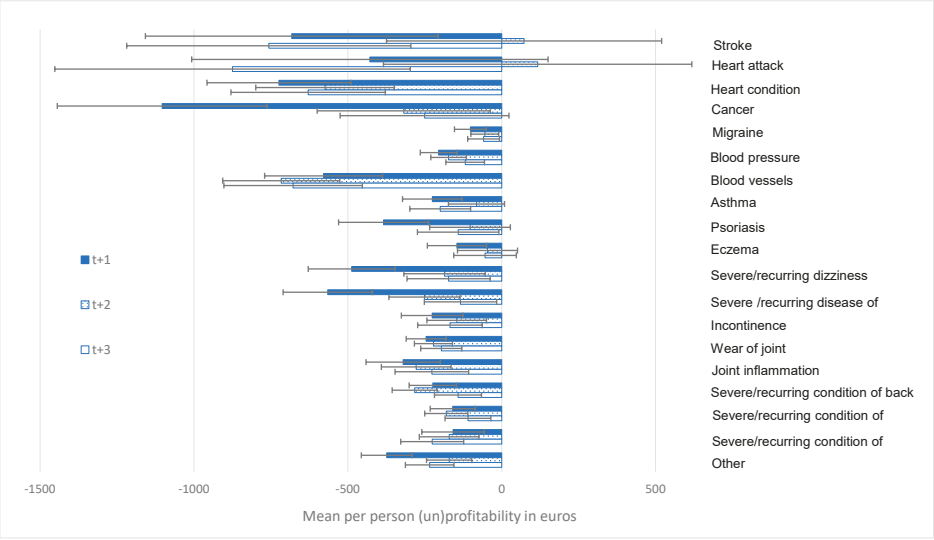


Note: (Un)profitability is defined as predicted spending minus actual spending. The 95% confidence intervals are calculated as $\bar{e_g} \pm 1.96 * \frac{\sigma_g}{\sqrt{n_g}}$

Figure 3.4 presents the (un)profitability for year t+1, t+2 and t+3 for 19 specific chronic illnesses that health survey respondents reported to have suffered from in the past 12 months. All these groups except one (heart attack) come with a statistically significant loss in year t+1, which for most groups persists in the years thereafter. Exceptions are stroke and heart attack which turn out to be profitable in

t+2 (but not statistically significant), as well as cancer, asthma, psoriasis, and eczema (statistically insignificant loss in t+2 and/or t+3). For most groups the unprofitability is smaller in the second and third year than in the first year, although these reductions may be driven by chance (as indicated by the often-overlapping confidence intervals).

Figure 3.4. (Un)profitability in euros after risk equalization for years t+1, t+2 and t+3 for individuals who in year t reported to have suffered from specific chronic conditions in the past 12 months



Note: (Un)profitability is defined as predicted spending minus actual spending. The 95% confidence intervals are calculated as $\bar{e_g} \pm 1.96 * \frac{\sigma_g}{\sqrt{n_g}}$

3.4 DISCUSSION

In many countries with regulated health insurance markets, health insurers are confronted with risk selection incentives towards specific subgroups. So far, studies have analyzed group-level (un)profitability for *one* contract period (e.g. Van Kleef et al. 2017a; Van Kleef et al. 2019; Bauhoff 2012; Buchner et al. 2013; Carey 2017a; Ellis et al. 2013b; is et al. 2017; Geruso et al. 2019; McGuire et al. 2013; McGuire et al. 2014; Shmueli & Nissan-Engelcin 2013; Withagen-Koster et al. 2018). However, a multiple contract period perspective might be more relevant as consumers do not necessarily switch health plans at the end of every contract period due to switching barriers. In this paper we have shown that when group-level (un)profitability is analyzed over multiple contract periods, insurers are persistently under- and overcompensated for chronically ill and healthy subgroups, respectively. This suggests that the incentives

for risk selection might be even stronger than could be assumed based on the results of previous research.

In our empirical analysis we simulated (un)profitability under sophisticated risk equalization for selective groups over multiple contract periods. More specifically, we identified 25 groups in data derived from a health survey conducted in year t, which we merged with administrative data for three consecutive follow-up years. For each of these groups we determined the mean actual spending, mean predicted spending and mean (un)profitability in each of these years. The results show that actual and predicted spending remain quite stable for the chronically ill subgroups. For the group of individuals who reported no chronic condition, there is a small but significant increase in both actual and predicted spending. Despite the recalibration of mean spending on a population level between the three years, we expected to find an increase in both spending and risk-equalization payments (i.e. predicted spending) as we expected a larger increase within specific (chronically ill) subgroups compared to the population. However, for most selective groups we found no statistically significant change in either. Furthermore, we found persistence in (un) profitability over time for almost all researched groups. More specifically, most chronically ill subgroups come with a persistent loss, while the group with no chronic condition is persistently profitable.

Persistent unprofitability of chronically ill could incentivize insurers to not be explicitly responsive to the preference of these groups and vice versa for healthy groups. A health insurer that is explicitly responsive to the preferences of unprofitable chronically ill, can expect to attract many chronically ill. Due to switching barriers, not all these chronically ill are likely to switch to another insurance plan next years, implying that the insurer remains with the unprofitable chronically ill for multiple contract periods. This is supported by recent research, which has indicated that health insurers in the Netherlands are reluctant to actively invest in the quality of health care for specific groups of chronically ill (KPMG 2014; 2020), due to the unprofitability of these groups (e.g. Van Kleef et al. 2019; Withagen-Koster et al. 2018; Van Kleef et al. 2013b). This strongly undermines the objectives of regulated competition in social health insurance. Elimination of predictable profits and losses is therefore crucial.

These findings for the Dutch setting (with one of the most sophisticated risk-equalization models in the world), are also relevant to other health insurance markets with a consumer choice of health plan, premium regulation and risk equalization. The reason is two-fold. First, like the Dutch risk equalization model, other risk-

equalization models too are known to undercompensate insurers for chronically ill people (see, for instance Bauhoff 2012; Buchner et al. 2013; Carey 2017a; Ellis et al. 2013b; Ellis et al. 2017; Geruso et al. 2019; McGuire et al. 2013; Shmueli & Nissan-Engelcin 2013). Second, in any individual health insurance market, consumers will experience switching barriers. As a result, specific groups might stay with the same health plan for multiple contract periods, especially when differences in premiums and plan value are limited. Reducing predictable variation in (un)profitability will then be important to mitigate risk-selection incentives and meet the objectives of regulated competition. To what extent a multiple year perspective is relevant for the evaluation of risk selection incentives depends on the switching behavior of consumers in the concerning market. Research so far has shown that chronically ill are less inclined to switch health plans compared to healthy individuals, which can be (partly) attributed to the higher switching barriers for chronically ill (Atherly et al. 2020; Boonen et al. 2016; Van der Schors et al. 2020). However, more research into switching behavior of consumers across markets is necessary.

Risk-selection incentives will be mitigated when the predictable profits/losses for groups of (chronically ill) individuals are reduced. Regulators have several options to achieve this. A first option is adding new risk adjusters to the risk-equalization model and/or refining existing risk-adjusters such that chronically ill are better identified in the model. If this is not possible due to for instance a lack of adequate data, 'modification of payment weights' is a promising alternative option. While payment weights for risk adjusters are typically estimated by an ordinary least squares (OLS) regression (as also done in this paper), recent work has shown that in terms of predictable profits and losses for groups of interest, alternative sets of payment weights might outperform those generated by OLS (Glazer & McGuire 2000; Withagen-Koster et al. 2020; Van Kleef et al. 2017b). A third option to reduce predictable losses is risk sharing, in which the regulator compensates health insurers retrospectively for (some) of the incurred spending. Risk sharing can take several forms, but usually comes with a trade-off between selection and efficiency (McGuire & Van Kleef 2018c). However, recent research has shown that some forms of risk sharing, like residual based reinsurance, can maintain adequate incentives for efficiency while reducing the incentive for risk selection (Van Kleef & Van Vliet 2022; McGuire et al. 2020). A final option is to allow some degree of premium differentiation. Health insurers can then price (a part of) the remaining predictable spending variation that the regulator can or will not compensate for through risk equalization. While this could stimulate insurers to become more responsive to the preferences of insured, it might also endanger the affordability of health insurance, especially for chronically ill.

All these options come with a tradeoff for the regulator. More research is needed on how these options can be optimized and/or combined in order to mitigate selection incentives while maintaining individual affordability of insurance coverage and incentives for efficiency.

3.5 CONCLUSION

Within regulated health insurance markets, health insurers are confronted with risk selection incentives. We have shown that group-level (un)profitability persists over multiple contract periods, in which chronically ill are persistently unprofitable and healthy subgroups are persistently profitable. Due to the persistent (un)profitability of chronically ill and healthy subgroups, health insurers possibly experience even greater risk selection incentives towards these groups than initially thought. Further mitigating these risk selection incentives is therefore important. This can be achieved in several ways, such as adding new or refining existing risk adjusters, the modification of payment weights, risk sharing or to allow some degree of premium differentiation. However, these options pose trade-offs to the regulator. More research is needed on how these options can be combined or optimized to further mitigate these risk selection incentives while safeguarding public goals like affordability.

APPENDIX A. Overview of the Dutch risk equalization models of 2016, 2017 and 2018

Table A.3.1. Risk adjusters and definitions of the three risk equalization models of 2016, 2017 and 2018 in the Netherlands

	Risk equalization model		
Risk adjusters	2016	2017	2018
Age / Gender[a]	Total of 40 age/gender classes.		Total of 42 age/gender classes. The classes for newborns have been split into being born last year (but not yet 1 year old) or the current year.
PCGs[a]	Total of 31 classes, including no PCG.	Total of 34 classes, including no PCG. Extremely high cost groups for medication has been added.	
(primary)DCGs[a]	Total of 16 classes, including no DCG.	Total of 16 classes, including nog DCG. One class has been split and was moved to another class. However, the underlying information for this risk adjuster did not change.	Total of 16 classes, including no DCG. Equal to 2017, except some extra diagnosis have been used to classify people.
Secondary DCGs	-	-	Total of 8 classes, including no secondary DCG. Added to better include co-morbidity.
DMECGs[a]	Total of 5 classes, including no DMECG.	Total of 11 classes, including no DMECG. Two classes have been removed and eight new classes have been added. To see the exact changes see Cattel et al.[b]	
MHCGs[a]	Total of 7 classes, including no MHCG.	Total of 8 classes, including no MHCG. One class has been split into two new classes.	Total of 9 classes, including no MHCG. One class for multiple year low costs has been added.
Source of Income[a]	Total of 24 classes of source of income interacted with age.		Total of 25 classes of source of income interacted with age. An extra class for highly educated has been added.
Region[a]	Total of 10 classes.		

Table A.3.1. Risk adjusters and definitions of the three risk equalization models of 2016, 2017 and 2018 in the Netherlands (*continued*)

	Risk equalization model		
Risk adjusters	2016	2017	2018
Socioeconomic status[a]	Total of 17 classes of social economic status interacted with age.	Total of 12 classes of social economic status interacted with age. The group who lives in an institutional home was separately classified. A class has been included in the new risk adjuster persons per street address and these individuals are classified with a low social economic status.	
Persons per street address	-	Total of 13 classes of persons per street address interacted with age.	
Physiotherapy spending in prior year[a]	Total of 2 classes, including no spending in the prior year.	-	-
PDCGs	-	Total of 5 classes, including no PDCG. Based on diagnosis information instead of spending.	
Home care spending in prior year[a]	Total of 5 classes, including no spending in the prior year.	Total of 7 classes, including no spending in the prior year. Two classes with new costs thresholds have been added.	Total of 8 classes, including no spending in the prior year. One class has been split based on whether they are an adult or a minor.
Geriatric rehabilitation care spending in prior year	Total of 2 classes, including no spending in the prior year.		-
Yes/no morbidity in interaction with age	Total of 4 classes.		-

[a] These risk-adjusters were included in the sophisticated risk equalization model in our simulations.

[b] Cattel, D., Eijkenaar, F., van Kleef, R.C., van Vliet, R.C.J.A.: Onderzoek risicoverevening 2017: Overall Toets. Rotterdam: iBMG/ Erasmus Universiteit Rotterdam (2016)

Note: This table is based on information from Eijkenaar, F., van Kleef, R.C., van Veen, S.H.C.M., van Vliet, R.C.J.A. 2015. Onderzoek risicoverevening 2016: Berekening normbedragen. Rotterdam: iBMG/ Erasmus Universiteit Rotterdam; Cattel, D., Eijkenaar, F., van Kleef, R.C., van Vliet, R.C.J.A. 2016. Onderzoek risicoverevening 2017: Berekening normbedragen. Rotterdam: iBMG/ Erasmus Universiteit Rotterdam and Cattel, D., Eijkenaar, F., van Kleef, R.C., van Vliet, R.C.J.A., Withagen-Koster, A.A. 2017. Onderzoek risicoverevening 2018: Berekening normbedragen. Rotterdam: iBMG/ Erasmus Universiteit Rotterdam.

APPENDIX B. Descriptive statistics

This Appendix compares the population frequencies of selected risk-adjuster variables in the administrative data for the three years. Any differences in definitions of spending or risk adjusters leading to large changes in population frequencies can have consequences for the outcomes of our simulations. Table B.3.1 shows descriptive statistics for individuals aged 18 years or older in each of the three years. The frequencies match relatively well. The differences in frequencies among the years are small and unlikely to have a notable impact on our results. The small differences in population frequencies across the three years can mainly be explained by differences in underlying definitions of specific risk-adjuster variables (see Appendix A). Specifically, the difference regarding classification in a DCG is due to changes in the underlying diagnosis registration system, which lead to more individuals being flagged by a DCG. The increase in population frequency of the DMECG risk adjuster is caused by an extension of this risk adjuster with a series of medical devices, resulting in more individuals being classified in a DMECG. A similar explanation holds for the risk adjuster 'physiotherapy spending in the previous year', for which the population frequency decreases from t+1 to t+2: as of t+2, individuals are classified in this risk adjuster based on diagnosis information instead of on the basis of yes/no spending (since not all diagnoses lead to classification, fewer individuals are being flagged by this risk adjuster). However, we do not expect these differences to have a notable impact on our results.

Table B.3.2 provides the index of the frequencies for individuals who in year t reported to have (not) suffered from at least one of the 23 chronic conditions (ever or in the past 12 months, depending on the condition) and for individuals who in year t reported to have ever suffered from diabetes, a stroke, a heart attack or cancer. The indices show a small decrease over time. As shown in Appendix C, this also holds for each of the 19 specific chronic illnesses which respondents reported to have suffered from in the past 12 months. This decrease over time can be caused by death, migration or unsuccessful merging of the health survey data with the relevant year of the administrative data.

Table B.3.1. Mean somatic spending and population frequencies for selected risk-adjuster variables of the adult population in the administrative data in t+1, t+2 and t+3

	Population in year t+1	Population in year t+2	Population in year t+3
N	13,136,338	13,196,124	13,278,671
Mean spending in euros per year	2414	2516[a]	2582[a]
Man, 19-34 year	12.9%	12.9%	12.9%
Man, 35-44 year	8.5%	8.2%	7.9%
Man, 45-54 year	9.6%	9.5%	9.5%
Man, 55-64 year	8.2%	8.2%	8.2%
Man, 65 year and older	9.8%	10.1%	10.4%
Woman, 19-34 year	12.6%	12.8%	12.7%
Woman, 35-44 year	8.6%	8.3%	8.0%
Woman, 45-54 year	9.5%	9.5%	9.5%
Woman, 55-64 year	8.2%	8.2%	8.3%
Woman, 65 year and older	12.0%	12.3%	12.5%
Pharmacy-based cost groups	23.6%	23.8%	24.1%
Diagnosis-based cost groups	11.3%	13.8%	12.1%
Multiple year high cost groups	6.9%	7.0%	7.0%
Durable medical equipment cost groups	1.1%	4.1%	4.2%
Physiotherapy	2.5%	1.2%	1.2%
Home care spending in the previous year	2.6%	3.0%	2.5%

[a] Mean spending for the adult population before recalibrating the spending levels in year t+2 and year t+3 to the spending level of year t+1 (i.e. 2013).

Table B.3.2. Index of the frequencies in year t+1, t+2 and t+3 for individuals who in year t reported to have (not) suffered from a chronic condition and for individuals who in year t reported to have ever suffered from cancer, a heart attack, a stroke or cancer

Groups identified in year t	Year	Frequency index
No chronic condition	t+1	100 (34% of population in year t+1)
	t+2	99
	t+3	98
At least one chronic condition	t+1	100 (53% of population in year t+1)
	t+2	98
	t+3	97
Diabetes	t+1	100 (6% of population in year t+1)
	t+2	97
	t+3	94
Stroke	t+1	100 (3% of population in year t+1)
	t+2	95
	t+3	89
Heart attack	t+1	100 (3% of population in year t+1)
	t+2	96
	t+3	90
Cancer	t+1	100 (6% of population in year t+1)
	t+2	95
	t+3	91

APPENDIX C. Frequency index, average spending and average predicted spending for years t+1, t+2 and t+3 for individuals who in year t reported to have suffered from a chronic illness in the past 12 months

Table C.3.1. Indices for the frequency of individuals, mean spending and mean predicted spending by the risk equalization model in year t+1, t+2 and t+3 for individuals who in year t reported to have suffered from a specific chronic condition in the past 12 months

Groups	Year	Frequency index	Mean per person spending	Mean per person predicted spending by risk equalization
Stroke	t+1	100 (0.5% of population in year t+1)	9071	8389
	t+2	94	8199	8272
	t+3	87	8754	7997
Heart attack	t+1	100 (0.4% of population in year t+1)	8483	8055
	t+2	95	7660	7777
	t+3	92	8274	7400
Heart condition	t+1	100 (2% of population in year t+1)	8897	8173
	t+2	95	8537	7964
	t+3	90	8500	7871
Cancer	t+1	100 (2% of population in year t+1)	10448	9345
	t+2	92	8463	8145
	t+3	87	7868	7617
Migraine	t+1	100 (15% of population in year t+1)	2301	2200
	t+2	99	2285	2230
	t+3	98	2332	2273
Blood pressure	t+1	100 (16% of population in year t+1)	4421	4216
	t+2	98	4442	4270
	t+3	96	4508	4390
Blood vessels	t+1	100 (3% of population in year t+1)	7585	7007
	t+2	95	7708	6992
	t+3	91	7489	6812
Asthma	t+1	100 (8% of population in year t+1)	4721	4495
	t+2	98	4577	4495
	t+3	95	4734	4534
Psoriasis	t+1	100 (3% of total population in year t+1)	3771	3387
	t+2	98	3611	3508
	t+3	97	3758	3617

Table C.3.1. Indices for the frequency of individuals, mean spending and mean predicted spending by the risk equalization model in year t+1, t+2 and t+3 for individuals who in year t reported to have suffered from a specific chronic condition in the past 12 months (*continued*)

Groups	Year	Frequency index	Mean per person spending	Mean per person predicted spending by risk equalization
Eczema	t+1	100 (5% of population in year t+1)	2783	2638
	t+2	99	2738	2692
	t+3	97	2735	2681
Severe/ recurring dizziness	t+1	100 (4% of population in year t+1)	5804	5317
	t+2	96	5447	5261
	t+3	92	5303	5130
Severe/ recurring disease of intestines	t+1	100 (4% of population in year t+1)	5272	4707
	t+2	97	4914	4663
	t+3	95	4744	4611
Incontinence	t+1	100 (6% of population in year t+1)	5617	5391
	t+2	96	5473	5326
	t+3	93	5440	5272
Wear of joint	t+1	100 (13% of population in year t+1)	4858	4612
	t+2	97	4893	4671
	t+3	95	4960	4763
Joint inflammation	t+1	100 (5% of population in year t+1)	5772	5452
	t+2	97	5872	5594
	t+3	95	5898	5671
Severe/recurring condition of back	t+1	100 (10% of population in year t+1)	4007	3783
	t+2	98	4110	3827
	t+3	97	4039	3897
Severe/ recurring condition of neck	t+1	100 (9% of population in year t+1)	3722	3563
	t+2	99	3777	3597
	t+3	97	3792	3682
Severe/ recurring condition of elbow	t+1	100 (6% of population in year t+1)	4374	4216
	t+2	98	4404	4234
	t+3	96	4498	4272
Other	t+1	100 (14% of population in year t+1)	4877	4504
	t+2	98	4572	4402
	t+3	95	4558	4324

4

Selection incentives in the Dutch basic health insurance: To what extent does end-of-life spending contribute to predictable profits and losses for selective groups?

With R.C. van Kleef & F. Eijkenaar

Based on: Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. (2022). Selection incentives in the Dutch basic health insurance: to what extent does end-of-life spending contribute to predictable profits and losses for selective groups? Medical Care Research and Review. https://doi.org/10.1177/10775587221099731.

ABSTRACT

Existing risk-equalization models in individual health insurance markets with premium-rate restrictions do not completely compensate insurers for predictable profits/losses, confronting insurers with risk selection incentives. To guide further improvement of risk-equalization models it is important to obtain insight into the drivers of remaining predictable profits/losses. This paper studies a specific potential driver: end-of-life spending (defined here as spending in the last 1-5 years of life). Using administrative (N=16.9m) and health survey (N=384k) data from the Netherlands, we examine the extent to which end-of-life spending contributes to predictable profits/losses for selective groups. We do so by simulating the predictable profits/losses for these groups *with* and *without* end-of-life spending while correcting for the overall spending difference between these two situations. Our main finding is that – even under a sophisticated risk-equalization model – end-of-life spending can contribute to predictable losses for specific chronic conditions.

4.1 INTRODUCTION

Social health insurance markets with open enrollment and premium-rate restrictions typically rely on risk equalization to mitigate incentives for risk selection. An example of such a health insurance market can be found in the Netherlands, where regulated competition has been introduced to improve efficiency of the healthcare system while safeguarding individual affordability of basic coverage and accessibility of care. In this system risk equalization is an important regulatory measure, which compensates insurers for predictable variation in healthcare spending among consumers (Van de Ven & Ellis 2000; Van de Ven et al. 2013).

Although the Dutch risk-equalization model is one of the most sophisticated in the world, it does not completely compensate for predictable spending variation. Since insurers are not allowed to risk rate their premiums, this results in predictable profits and losses on specific subgroups of consumers. For example, recent research using diagnostic information from general practitioners has shown that Dutch insurers face a predictable loss of approximately 85 euros per person per year on the group with at least one chronic condition (approximately 52 percent of the population) and a predictable profit of 90 euros per person per year on the complementary group without a chronic condition (Van Kleef et al. 2018b). This incentivizes insurers to attract healthy people (e.g. by selective marketing towards these groups) and to deter the chronically ill (e.g. by quality skimping). Such risk selection can distort both the efficiency of the health insurance market (e.g., via quality distortions) and jeopardize fairness of healthcare financing (e.g. when predictably profitable and predictably unprofitable people sort into different insurance plans, which threatens the level playing field and might result in selection-driven premium variation across plans) (Glazer & McGuire 2000; Van Kleef et al. 2019; Van de Ven et al. 2015). While the presence of risk selection is very difficult to demonstrate, Dutch health insurers have stated that they are reluctant to actively invest in health care for specific groups of chronically ill that remain to be undercompensated by the risk equalization model (KPMG 2014; KPMG 2020). This illustrates the importance of reducing the remaining predictable profits and losses. To guide further improvement of risk equalization systems it is important to obtain insight into the drivers of existing predictable profits and losses.

One possible factor that may contribute to the existing predictable profits and losses is end-of-life spending (defined here as spending in the last 1-5 years of life). On average, spending tends to be higher for individuals who are in the last phase of their life compared to those who are not (Polder et al. 2006; Van Vliet & Lamers 1998). In

addition, a significant share of total healthcare spending in a year can be attributed to end-of-life spending (Shmuelli et al. 2010; Stooker et al. 2001). Moreover, Van Vliet & Lamers (1998) have shown that on average, healthcare spending of decedents starts to increase six years prior to death. To the extent that individuals who are in the last phase of their life are overrepresented in the subgroups used for evaluating selection incentives and the risk-equalization model does not adequately compensate for their high spending, end-of-life spending might significantly contribute to the remaining predictable profits and losses for these subgroups. Although the Dutch risk-equalization model includes many risk adjusters that compensate for high spending related to age, gender and health, this model does not include a risk adjuster that explicitly flags people who are near the end of their life (Van Kleef et al. 2018a). Therefore, end-of-life spending might indeed contribute to remaining predictable profits/losses on selective groups. On the other hand, the end-of-life stage is likely to correlate with existing morbidity indicators in the risk equalization model as a large share of the ex-post spending on the deceased can be explained by ex-ante spending on the sick (Einav et al. 2018). Therefore, the morbidity-based risk adjusters in the risk-equalization model may already compensate for a large share of end-of-life spending (Van Vliet & Lamers 1998).

The aim of this paper is to gain insight in the extent to which end-of-life spending contributes to existing group-level predictable profits and losses under sophisticated risk equalization. This insight will be helpful in guiding further improvements of the Dutch risk-equalization scheme. If we find a significant contribution of end-of-life spending to predictable profits and losses, improvement of the Dutch risk equalization scheme should (also) focus on end-of-life spending, e.g., by refining existing morbidity indicators or by applying some form of risk sharing. If we do not find a significant contribution, apparently other factors drive predictable profits and losses, and further research will be necessary to identify these factors. We use administrative data from the Netherlands (N≈16.9m) on both actual and predicted spending for 2013, and on whether someone has died during the period 2013-2017. We merge these data with health survey data from 2012 (N≈384k) and calculate predictable profits and losses in 2013 for selective subgroups of healthy and chronically ill individuals.

4.1.1 New contribution

An important difference between our paper and previous papers focusing on end-of-life spending (Van Vliet & Lamers 1998; Einav et al. 2018) is that rather than focusing on the predictability of end-of-life spending itself, we primarily focus on the *contribution* of end-of-life spending to selection incentives regarding selective

groups of individuals. The underlying assumption is that insurers are likely and able to risk select based on health (e.g. yes/no chronic condition) rather than on yes/no being near the end of life. While 'health' can be known before the start of a contract period and thus be acted upon by insurers (e.g. in terms of marketing and design of health plans) and/or consumers (i.e. by choosing a certain plan), 'being near the end of life' is hard to predict and thus difficult to act upon (Einav et al. 2018). To our knowledge, this is the first paper that explicitly examines the extent to which end-of-life spending contributes to predictable profits and losses for subgroups that are vulnerable to risk selection. Since we use the actual risk-equalization data from the Netherlands, our empirical findings are directly relevant for that specific context. The international relevance of our paper is to be found in the conclusion that – even under a sophisticated risk-equalization model as the one applied in the Netherlands – end-of-life spending can indeed contribute to the predictable losses for specific groups. What this contribution looks like in other settings is an empirical question and depends on specific contextual factors such as the benefits package, the characteristics of the population and the features of the risk-equalization model in place. Our methodology can be useful for identifying the contribution of end-of-life spending to predictable profits and losses in other countries/settings.

This paper is structured as follows. The next section describes the institutional setting, followed by the data and methods used. Then the results are reported. The last section summarizes and discusses the main findings.

4.2 THE DUTCH HEALTH INSURANCE SYSTEM

The Dutch health insurance system consists of three components: a mandatory public insurance scheme for long-term care, a mandatory basic health insurance scheme providing coverage for curative care (e.g. primary care, pharmaceutical care, inpatient and outpatient hospital care and mental health care), and a voluntary supplementary health insurance covering services not covered by the two mandatory schemes. This paper focuses on the basic health insurance system for curative care and all spending covered under that system, which operates on the basis of regulated competition among private insurers (Van Kleef et al. 2018a). In this system, competition among insurers is driven by a free consumer choice of insurance plan. Insurers have some flexibility regarding provider network and coverage of out-of-network spending, resulting in competition among providers of care. Relevant regulatory measures include an annual open enrollment requirement, community-

rating per health plan, a standardized benefits package, an individual mandate to buy health insurance and risk equalization.

The Dutch risk-equalization model of 2016, which is the focus of this paper and has since 2016 undergone only relatively small changes, contains a broad set of socio-demographic risk adjusters (e.g., age interacted with gender, socioeconomic status and source of income interacted with age) as well as seven morbidity-based risk adjusters. Each risk adjuster consists of multiple risk classes, 162 in total. Payment weights for risk classes are estimated with a multivariate least-squares regression (of spending on risk classes) using data from a prior period (which has been made representative for payment year t in terms of benefits package, projected spending and composition of the population). This results in a prediction model that allows for calculating individual-level spending in euros, which forms the basis for the risk-equalization payments. More specifically, the risk-equalization payment for enrollee *i* is calculated as *i*'s predicted spending minus a fixed amount Y. The value for Y is determined by the Minister of Health and reflects the amount of spending that has to be financed via the out-of-pocket premium. The community-rated out-of-pocket premium reflects Y as well as the relative (in)efficiency of insurers, thereby creating price competition. From the perspective of insurers, part of their revenues (about 50%) comes from the risk-equalization fund (which itself is financed by earmarked income-related contributions) and the other part (also 50% in total) comes from out-of-pocket premiums. Risk-equalization transfers are executed by the Dutch Healthcare Institute.

The morbidity-based risk adjusters in the Dutch model include Pharmacy-based Cost Groups (PCGs), Diagnosis-based Cost Groups (DCGs), Multiple year High-Cost Groups (MHCGs), Durable Medical Equipment Cost Groups (DMECGs), physiotherapy spending in the prior year, geriatric rehabilitation care spending in the prior year and home care spending in the prior year. Each of these morbidity adjusters is 'prospective, which means that these indicators are based on information from a prior period. For example, the PCG-adjuster consists of 33 classes based on individuals' medication use in the *prior* year. To be flagged by a PCG, individuals must pass a predetermined defined daily dose (DDD) threshold of the relevant medication. The DCG-adjuster comprises 15 classes in which individuals can be classified based on specific diagnoses from inpatient and outpatient hospital treatment in the prior year. The MHCG-adjuster contains 7 classes based on the level of spending for curative somatic care in the three prior years. The assumption is that individuals with multiple-year high costs most likely suffer from a chronic condition. Individuals can be flagged by one of the 4 classes of the DMECG-adjuster based on the use of specific

durable medical equipment in the prior year. The remaining three morbidity-based risk adjusters are all based on prior-year spending on specific types of care, i.e. physiotherapy, geriatric rehabilitation care, and home care. For more information about the Dutch risk-equalization system, we refer to Van Kleef et al. (2018a).

4.3 DATA & METHODS

4.3.1 Data

To examine the contribution of end-of-life spending to group-level profits and losses, we merged three datasets using an anonymized individual-level identification key. The first dataset includes administrative data on individual-level spending and risk adjusters for all Dutch citizens with a basic health insurance in 2013 (N≈16.9 million). These data allow us to replicate the Dutch risk equalization model.

The second dataset contains information on whether someone has died in the period 2013-2017 and, if so, in which year. This information enables us to identify people in our 2013-data who are near the end of their life. Given our data, we take into account five definitions of 'being near the end of life': deceased within 1 year (i.e. died in 2013), deceased within 2 years (i.e. died in 2013-2014), deceased within 3 years (i.e. died in 2013-2015), deceased within 4 years (i.e. died in 2013-2016) and deceased within 5 years (i.e. died in 2013-2017).[7]

The third dataset comes from a health survey conducted in 2012 (N≈384k)[8]. This dataset contains information on self-reported chronic conditions by individuals aged 19 years or older on September 1, 2012 (Public Health Monitor, 2012). We use these data to define subgroups that are potential targets of risk selection by insurers (e.g. by selective advertising and insurance plan design). These groups have been extensively analyzed in previous studies and are considered relevant when it comes to the evaluation of the Dutch risk-equalization model in terms of selection incentives (Van Kleef et al. 2013a; Van Kleef et al. 2013b; Van Kleef et al. 2017a; Van Kleef et al. 2019; Withagen-Koster et al. 2020).

Before conducting the evaluation of predictable profits/losses, we rebalanced the health survey sample using a raking procedure. Although the unbalanced sample

7 Van Vliet & Lamers (1998) showed that the above-average spending of decedents starts increasing six years prior to death, with the increase becoming steeper as people are closer to death.

8 Public Health Monitor (2012) of the Community Health Services, Statistics Netherlands and the National Institute for Public Health and the Environment.

is already quite representative for the population (Van Kleef et al. 2017a), this procedure enabled us to further improve upon this. The raking procedure is an iterative process that generates a weight for every record in our dataset using a set of key variables that are present in both the survey sample and the total population. Application of these weights ensures that the frequencies of these variables in the sample are similar to those in the population. (Izrael et al. 2009). The set of key variables includes all risk adjuster classes of the Dutch risk-equalization model of 2016, as well as 18 quantiles of mean curative somatic spending and a proxy for whether someone had died in 2013.[9] For a more detailed description of the rebalancing procedure as well as results on the sample's representativeness before and after rebalancing, see Withagen-Koster et al. (2020).

4.3.2 Methods

To determine the contribution of end-of-life spending to predictable profits and losses for selective subgroups we performed a simulation that consisted of seven steps. First, we merged the administrative data on 2013 spending and risk-adjuster flags with the data on the year of death (indicating whom in the administrative data of 2013 died in the period 2013-2017). In an explorative analysis, we examined the characteristics in 2013 of those who died in the period 2013-2017.

Second, we used the health survey data to define 25 selective subgroups that are potentially vulnerable to risk selection. More specifically, we identified nineteen specific chronic conditions that individuals in 2012 could report to have suffered from in the past twelve months, and four chronic conditions that individuals in 2012 could report to have suffered from ever in the past. In addition, we constructed two more general groups based on yes/no self-reported chronic condition (ever or in the past twelve months).

Third, as a baseline, we estimated the actual Dutch risk-equalization model of 2016 on the total population (N=16.9m). This baseline model will be referred to as M1. Under M1, the payment an insurer receives for a certain enrollee equals the predicted spending for that enrollee generated by the risk-equalization model.

Fourth, we supplemented M1 with a 100% cost-based compensation for people near the end of their life. We refer to this model as M2. More specifically, we applied the Dutch risk-equalization model of 2016 to the total population *excluding* those near

9 Information on whether someone had actually died was not available for the rebalancing procedure, so we had to resort to a proxy.

the end of their life. For individuals in the group on which the model was estimated, payments equal predicted spending generated by this model. For excluded individuals (i.e. those near end of life), we set the payment equal to actual spending (i.e. a 100% cost-based compensation). This scenario essentially implies 'carving out' end-of-life spending. Compared to M1, M2 is expected to reduce predictable profits and losses on the subgroups of step 2, because it is likely these subgroups include some decedents with end-of-life spending. This reduction, however, is not necessarily (fully) attributable to end-of-life spending and can thus not be interpreted as such. The reason is that *any* 'carving out' of spending is likely to reduce predictable profits and losses. For example: in the extreme situation in which *all* spending would be 'carved out' (i.e. 100% cost-based compensation for all individuals in the population), predictable profits and losses would drop to zero.

To determine whether the reduction in predictable profits and losses under M2 is indeed attributable to end-of-life spending, we cannot compare M2 to M1 as this would be an apples-to-oranges comparison due to the difference in overall mean spending. Therefore, we defined a third model (M3). This is the fifth step of our simulation analysis. Whereas M2 supplements the baseline model (M1) with a 100% cost-based compensation for people near the end of life, M3 supplements M1 with a form of *proportional* cost-based compensation for the *entire* population with positive spending. Specifically, under M3 – which functions as a counterfactual for M2 – we carve out the same proportion of total spending as under M2, but then across the entire population instead of only for certain groups. While under M2 we carve out *100%* of spending for *individuals near the end of life*, under M3 we carve out *X%* of spending for *all* individuals, with X being the share of end-of-life spending in total spending (defined as the sum of spending for people near the end of their life divided by the sum of spending in the total population. See also the last column of Table 4.2 below).

Sixth, for each of the 25 subgroups identified in step 2 and for each of the models M1-M3 described in steps 3-5, we calculated the mean per person profit/loss as the mean payment that insurers receive for a group minus the mean spending for that group.[10]

10 Because the risk-equalization model is estimated by OLS, mean spending equals mean predicted spending in the estimation data. However, this is not necessarily the case for samples drawn from these data. Therefore, after the rebalancing but before calculating the group-level predictable profits and losses, we corrected the mean spending under each model estimation in the survey sample such that the overall mean spending and the overall mean predicted spending under each model are equal. We did this by multiplying individual-level spending by a factor obtained from dividing the mean predicted spending by the mean spending. These factors are 1.00246 for the base model and 1.00575, 1.0062, 1.00787, 1.0067 and 1.00693, respectively each of the definitions of end-of-life.

Finally, for each group we tested whether the differences in the profits/losses between M2 and M3 are statistically significant using a paired t-test. If the profit/loss for a certain group under M2 (with spending being carved out for those near end of life) is statistically significantly different from that under M3 (with the same proportion of total spending being carved out as under M2, but then in the form of a percentage of spending for each and every individual in the population), we can conclude that end-of-life spending contributes to the existing profit/loss for that group under the baseline scenario (M1).

We repeated steps 4-7 for each of the five definitions of 'being near the end of life' described in the Data section.

4.4 RESULTS

This section presents the results of our simulations. First, some descriptive statistics of the administrative and health survey data are presented, followed by the characteristics of the deceased. Next, we provide information on the proportion of the population near the end of life in our subgroups. Lastly, we quantify the contribution of end-of-life spending to the predictable profits and losses on these groups.

4.4.1 Descriptive statistics

Table 4.1 shows descriptive statistics of the unbalanced and rebalanced survey sample and the population in the administrative data for individuals aged 19 years and older on September 1, 2012. After rebalancing, the statistics for the sample match those in the population very well.

Table 4.2 provides descriptive information on our five definitions of being near the end of life. Of the total population in 2013, 0.8% died in 2013. We find similar annual death rates for later years, leading to 4.3% of the 2013-population having died in the period 2013-2017. For those who died in 2013, spending amounts to 5.3% of total spending. This percentage increases to 20.1% after inclusion of those who died in the four years thereafter. Note that the percentages in the last column are used as a basis for the proportional cost-based compensations under M3 (i.e. they represent the proportion 'X' as described in step 5 in the Methods section).

Table 4.1. Mean curative somatic spending and population frequencies in 2013 for selected risk-adjuster variables in the unbalanced and rebalanced survey sample and the in terms of age corresponding population in the administrative data (19 years or older on September 1, 2012).

	Survey sample (unbalanced)	Survey sample (rebalanced)	Population
N	384,004	384,004	12,774,886
Mean spending in euros (sd)	3216 (8909)	2460 (7793)	2460 (8016)
Mean predicted spending (sd)	3247 (5179)	2460 (4558)	2460 (4554)
Man, 19-34 year	5.6%	11.8%	11.8%
Man, 35-44 year	4.8%	8.7%	8.7%
Man, 45-54 year	6.6%	9.8%	9.8%
Man, 55-64 year	7.6%	8.4%	8.4%
Man, 65 year and older	20.7%	10.1%	10.1%
Woman, 19-34 year	8.0%	11.8%	11.8%
Woman, 35-44 year	6.5%	8.8%	8.8%
Woman, 45-54 year	8.4%	9.8%	9.8%
Woman, 55-64 year	8.5%	8.4%	8.4%
Woman, 65 year and older	23.4%	12.4%	12.4%
PCGs	33.1%	24.1%	24.1%
DCGs	16.2%	11.5%	11.5%
MHCGs	10.4%	7.1%	7.1%
DMECGs	1.5%	1.1%	1.1%
Physiotherapy spending in the previous year	3.8%	2.6%	2.6%
Home care spending in the previous year	3.7%	2.6%	2.6%
Geriatric rehabilitation care spending in the previous year	0.3%	0.3%	0.3%

Note: PCGs are Pharmacy-based Cost Groups, DCGs are Diagnosis-based Cost Groups, MHCGs are Multiple year High-Cost Groups and DMECGs are Durable Medical Equipment Cost Groups.

Table 4.2. Deceased as percentage of total population in 2013 and end-of-life spending as percentage of total spending in 2013, for five definitions of being near the end of life.

Definitions of being near end of life	Deceased as % to total population in 2013	Spending of people near end of life as % of total spending in 2013
Deceased in 2013	0.8%	5.3%
Deceased in 2013-2014	1.6%	10.5%
Deceased in 2013-2015	2.5%	14.2%
Deceased in 2013-2016	3.4%	17.3%
Deceased in 2013-2017	4.3%	20.1%

4.4.2 Characteristics in year t of those being near the end of life (survey sample)

This section presents characteristics in 2013 for the following groups (conditional on the survey sample): those who died in 2013, those who died in 2014, those who died in 2015, those who died in 2016 and those who died in 2017. For each of these five groups, Table 4.3 shows the mean actual and predicted spending, the mean profit/loss, the percentage flagged by a morbidity-based risk adjuster, and the percentage with a self-reported chronic condition. For those who died in the period 2013-2017, average spending in 2013 is higher than for the average person in the sample. The same is true for predicted spending in 2013, indicating that the risk-equalization model explains some of the spending variation between people who are near the end of their life and those who are not. This is supported by the high percentage of individuals flagged by at least one morbidity-based risk adjuster.

Table 4.3 also shows a significant gap between predicted and actual spending for all groups. This gap is largest for those who are within one year from death (loss of 19,558 euros) and smallest for those who are 3 to 4 years from death (loss of 721 and 440, respectively). On average, spending in 2013 for those who died between 2013 and 2017 equals 15,879 euros while predicted spending in 2013 for this group equals 10,112 euros, implying an ex-post loss of 5,767 euros in 2013. Note that this ex-post loss is not necessarily predictable ex-ante.[11] However, to the extent that this ex-post loss is concentrated in our subgroups of interest, it might contribute to the predictable losses on these groups.

Table 4.3 further shows that people near the end of life are much more likely to be flagged by at least one morbidity-based risk adjuster than the average person in the sample (31.0%). The percentage with at least one morbidity flag is somewhat higher for those who are less than one year from death (84.3%) than those who are more than one year from death (between 79.2% and 74.7%). The percentage of individuals with at least one self-reported chronic condition is also much higher among those near the end of life (ranging from 74.2% to 80.7%, depending on the definition) than in the sample (53.3%). Moreover, as becomes clear from the bottom three rows, people who are near the end of their life suffer more often from multimorbidity. This makes sense since those who are very ill are more likely to die. This is also what we see when we analyze the relative frequency of decedents within each of the morbidity-based risk classes (as described in 'The Dutch health

11 In this study, our primary focus is not on the predictability of death itself, but on the extent to which end-of-life spending contributes to group-level predictable profits and losses.

insurance system' section). When we look at the relative frequency of those who died in 2013, we find the highest values for DCG 11 (17 percent) which includes people with cancer and spina bifida, for DCG 13 (11 percent) which includes people with aplastic anemia and users of home ventilation, for PCG 17 (10 percent) and 29 (13 percent) for Parkinson's disease and cancer medication respectively, and lastly for the highest class of the home care risk adjuster (11 percent), which includes individuals with spending in the top 0.25% for home care in the prior year. When we look at the relative frequency within each of the morbidity-based risk adjusters conditional on the group of people who died in the period 2013-2017, we find the highest values for the highest three classes of the home care risk adjuster (46 to 58 percent) which includes individuals with spending in the top 1.5, top 0.5 and top 0.25 percent of home care spending in the prior year, DCG 15 for people requiring renal dialysis and PCG 26 for kidney disease (both 46 percent).

Table 4.3. Mean (predicted) spending, mean ex-post profit/loss, percentage flagged by at least one morbidity-based risk-adjuster, and percentage with (at least) one or more self-reported chronic condition(s) in 2013, for the total survey sample and for those who died in the period 2013-2017 in that sample.

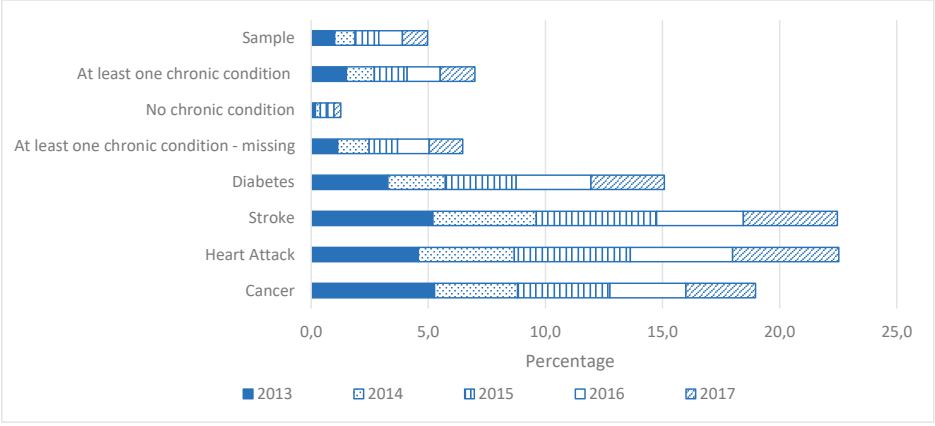
			Died in				
		Total sample	2013	2014	2015	2016	2017
Mean spending in euros (sd)		2460 (7793)	32408 (42498)	17436 (25854)	11388 (17421)	9480 (16391)	8684 (14774)
Mean predicted spending in euros (sd)		2460 (4558)	12850 (12802)	11496 (13306)	9210 (11329)	8759 (9689)	8244 (10256)
Mean ex-post profit/loss in euros		0	-19558	-5940	-2178	-721	-440
Flagged by at least one morbidity-based risk adjuster included in the risk-equalization model		31.0%	84.3%	79.2%	77.8%	76.1%	74.7%
At least one self-reported chronic condition (ever or in past 12 months)	Yes	53.3%	80.7%	72.9%	74.2%	74.5%	74.2%
	No	34.3%	6.2%	8.1%	10.8%	9.0%	9.2%
	Missing	12.4%	13.2%	19.0%	14.9%	16.6%	16.6%
Diabetes (ever)		5.8%	17.8%	16.3%	17.5%	18.5%	17.2%
Stroke (ever)		2.9%	15.9%	14.8%	14.9%	10.7%	11.0%
Heart Attack (ever)		3.0%	13.6%	14.1%	14.9%	13.1%	12.9%
Cancer (ever)		6.5%	34.3%	26.6%	25.8%	21.2%	18.4%
Number of self-reported chronic conditions (ever or in past 12 months)	One	24.6%	20.1%	24.1%	25.9%	25.1%	26.2%
	Two	13.1%	19.3%	20.3%	18.8%	19.8%	19.4%
	Three	6.8%	16.1%	14.4%	13.7%	14.7%	14.6%
	Four or more	8.7%	37.4%	31.2%	28.8%	29.7%	28.8%

Note. The 'number of self-reported conditions' is based on the conditions respondents could report to have suffered from ever in the past (i.e. the four conditions listed in this table) as well as the conditions respondent could report to have suffered from in the past 12 months (listed in table B.4.1)

4.4.3 Share of deceased in subgroups from the health survey

The previous section has shown that individuals with a morbidity-based risk adjuster (i.e. suffering from a chronic condition according to the risk equalization model) are overrepresented among those near the end of life. But to what extent are individuals near the end of life overrepresented among those who reported to suffer from a chronic condition in the health survey? Figure 4.1 answers this question for a selection of subgroups identifiable in the survey sample. As can be seen, 5 percent of the total sample has died after five years (2013-2017). In line with expectations, this percentage is indeed higher in the group 'at least one chronic condition' (7%) and lower for the group 'no chronic condition' (1.3%). The bottom four groups are specific chronic conditions that individuals could report to have ever suffered from in the past. The group 'cancer' has the highest death rate in the first year (5.3%), while after five years the cumulated rate is highest in the groups 'stroke' and 'heart attack' (both 22.5%).

Figure 4.1. Decedents in 2013, 2014, 2015, 2016 and 2017 as a percentage of the total survey sample in 2013 and of specific subgroups identifiable in that sample.



Note: 'Sample' refers to the whole health survey sample. The groups 'at least one chronic condition', 'no chronic condition' and 'at least one chronic condition-missing' are determined based on chronic conditions respondents could report to have ever suffered from or in the past 12 months. The specific groups 'diabetes', 'stroke', 'heart attack' and 'cancer' refer to chronic conditions people could report to have ever suffered from in the past.

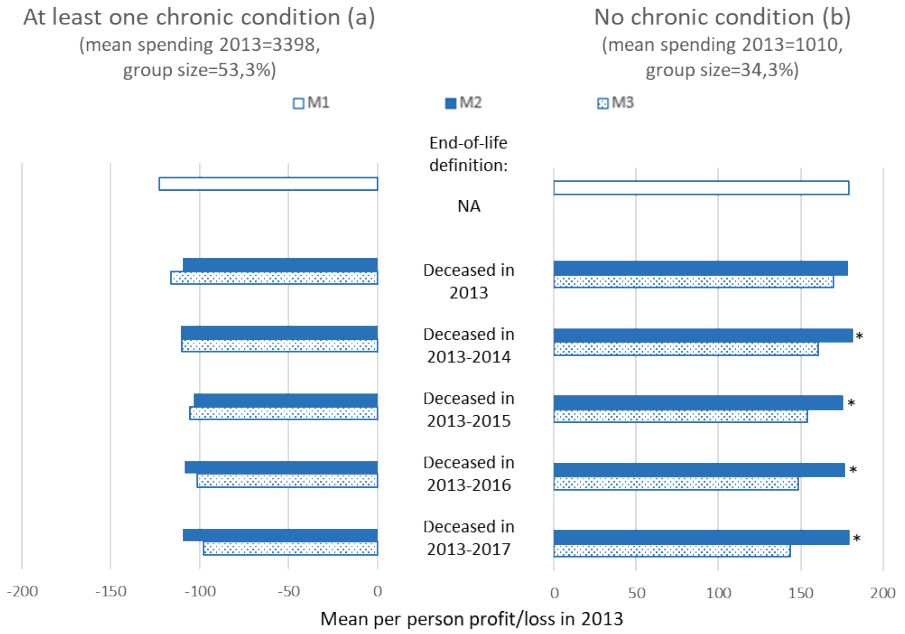
4.4.4 Contribution of end-of-life spending to group-level profits and losses

This section presents the mean per person profits/losses by subgroup under the three models (M1-M3) for our five definitions for 'end of life'. M1 replicates the Dutch risk-equalization model of 2016, M2 is M1 supplemented with 100% cost-based compensation for (spending of) people who are near the end of their life,

and M3 is M1 supplemented with a proportional cost-based compensation using the percentages in the right column of Table 4.2. Figure 4.2 shows the results for these models for two groups: those who in 2012 (t-1) reported to have suffered from at least one chronic condition (ever or in the past 12 months, panel a) and those who reported not to have suffered from a chronic condition (panel b). As expected, under M2 the predictable loss on the group with a chronic condition is lower than under M1, for each definition of being near the end of life. For end-of-life spending to have really contributed to the predictable loss on this group, the loss under M2 (filled bars) must be statistically significantly lower than the value under M3 (scattered bars). As this is not the case for any of the five definitions of end-of-life spending, we cannot conclude that end-of-life spending contributes to the predictable loss on this group. For the group without a chronic condition, we find that the predictable profit does not really reduce under M2 compared to M1. In an additional analysis (not shown here) we found that – for this specific group – actual spending and predicted spending roughly reduce to the same extent, thereby leaving the size of the gap (i.e. the predictable profit) intact. As a result, the predictable profit under M2 is higher than under M3. (Note that under M3 both actual spending and predicted spending decrease proportionally compared to M1, implying that because this groups is overpaid, the absolute reduction is higher for predicted spending than for actual spending. Consequently, the predictable profit for this group decreases under M3 compared to M1.)

We also examined the mean per person profit/loss under models M1-M3 for the group of individuals for whom the information 'yes/no self-reported chronic condition' is missing (see appendix A). The per person profit under M1 is much closer to zero compared to the groups 'at least one self-reported chronic condition' and 'no self-reported chronic condition', indicating that the group for which this information is missing is not a selective group. For this group we also do not find a statistically significant difference between M2 and M3 for any of the five definitions of end-of-life spending indicating that end-of-life spending does not contribute to the remaining profits.

Figure 4.2. Mean per person profit/loss in 2013 under M1-M3 for five definitions of being near the end of life for the group who in 2012 reported to suffer from at least one chronic condition and the group who in 2012 reported to suffer from no chronic condition.

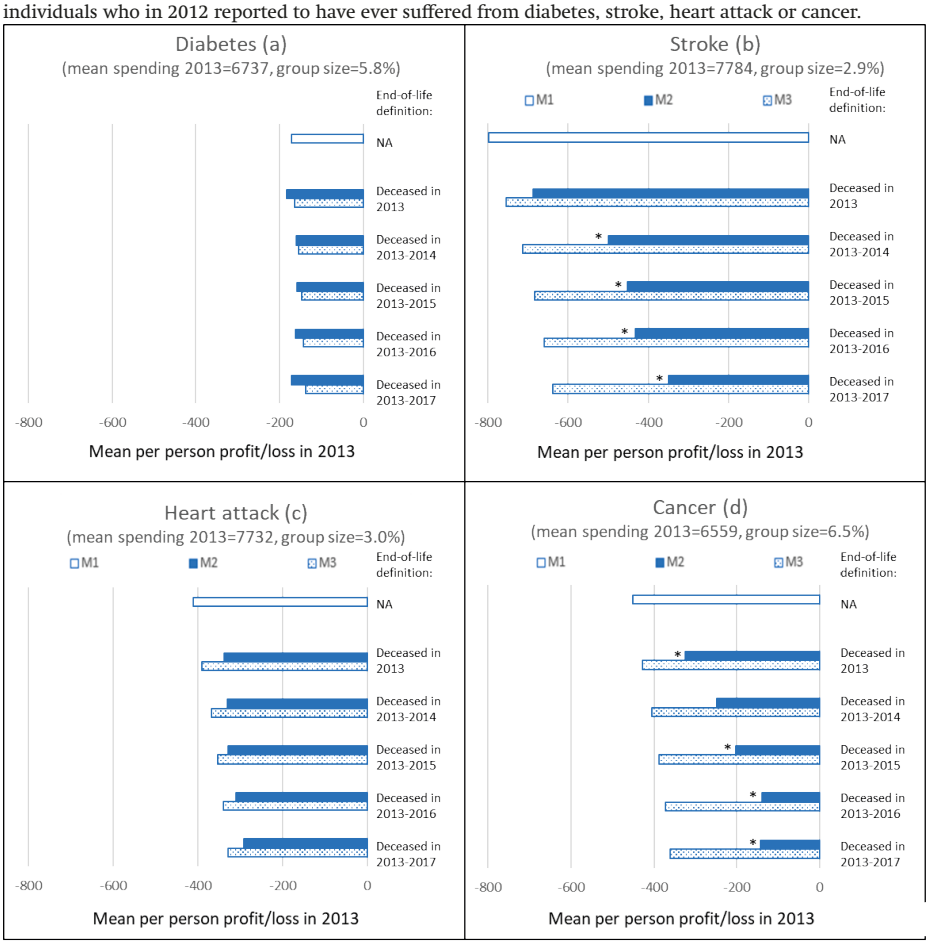


Note: Due to the group with missing values for yes/no self-reported chronic condition, the groups in panel a and b do not add up to 100 percent. See appendix A for the results on the group with missing values. M1 is the Dutch risk-equalization model 2016, M2 is M1 supplemented with 100% cost-based compensation for people who are near the end of life, and M3 is M1 supplemented with a proportional cost-based compensation using the percentages in the right column of Table 4.2. NA means not applicable. An asterisk (*) next to a set of bars indicates a statistically significant ($p<0.05$) difference between the profit/loss under M2 versus M3.

Figure 4.3 shows results for individuals who reported to have ever suffered from diabetes (panel a), stroke (panel b), heart attack (panel c) or cancer (panel d). For diabetes, M2 results in a slightly higher predictable loss than M1 under the definition 'deceased in 2013'. With the expansion of the end-of-life definition, a small decrease followed by a small increase can be observed, but none of these results are statistically significantly different from those under M3.

For stroke, the loss under M2 is smaller than under M1 for all definitions of end of life (panel b). Except for the definition 'deceased in 2013', the loss is also statistically significantly lower than under M3, suggesting that end-of-life spending contributes to this loss. Compared to M3, the loss is 30% to 45% lower, depending on the definition of being near end of life.

Figure 4.3. Mean per person loss in 2013 under M1-M3 for five definitions of being near the end of life for individuals who in 2012 reported to have ever suffered from diabetes, stroke, heart attack or cancer.



Note: M1 is the Dutch risk-equalization model 2016, M2 is M1 supplemented with 100% cost-based compensation for people who are near the end of life, and M3 is M1 supplemented with a proportional cost-based compensation using the percentages in the right column of Table 4.2. NA means not applicable. An asterisk (*) next to a set of bars indicates a statistically significant ($p<0.05$) difference between the profit/loss under M2 versus M3.

Panel c shows that for the group 'heart attack', the predictable loss under M2 keeps declining as the end-of-life definition encompasses more years before death. However, there is no statistically significant difference with the losses under M3 for any of the definitions.

Lastly, panel d shows a declining loss for the group 'cancer' under M2 compared to M1 for all definitions. The loss under M2 is statistically significantly lower than that

under M3 for al definitions of end of life. Compared to M3, this loss is approximately 40% to 60% lower, depending on the definition.

Appendix B shows the mean per person profit/loss under the three models M1-M3 and five definitions of being near end of life for 19 specific chronic conditions respondents could report to have suffered from in the past 12 months (instead of ever in the past, like in Figure 4.3). For 15 groups, the predictable losses under M2 are not significantly different from those under M3 for any of the definitions of being near the end of life. The four exceptions are the groups 'cancer', 'blood pressure' 'blood vessels' and 'joint inflammation'. In general, for these groups the difference in predictable loss between M3 and M2 is more likely to be statistically significant for the more comprehensive end-of-life definitions.

4.5 DISCUSSION

In this paper we determined the contribution of end-of-life spending to predictable profits and losses that insurers face for selective subgroups in the Dutch basic health insurance. In line with prior research (Van Vliet & Lamers 1998; Einav et al. 2018), our descriptive analyses show that decedents have higher spending on average and that spending is already higher up to five years prior to death. In addition, those who will die within 5 years are more likely to be flagged by a morbidity-based risk adjuster included in the risk-equalization model. Our descriptive analyses further show that on average people near the end of life do not only suffer more often from a chronic disease, but also suffer more often from multimorbidity.

To determine the contribution of end-of-life spending to predictable profits and losses on groups of interest, we simulated these profits and losses with and without end-of-life spending, while correcting for the overall difference in mean spending between these two situations. Our results show that end-of-life spending contributes to the predictable profits and losses for some groups, but not for others. Accounting for end-of-life spending significantly reduced the predictable losses for some groups, like individuals who indicated to have suffered from stroke (ever) or cancer (ever or in the past year), or a condition of the blood vessels (in the past year). Individuals with these chronic conditions might not always be flagged as such by the risk-equalization model, for instance because they do not cross the defined daily dose (DDD) threshold to be flagged by a PCG (although they might use the relevant medication), and/or because they did not have a hospital treatment in the prior year (and thus no DCG flag). This could explain why we find an effect for these groups.

To reduce selection incentives for these specific groups, better incorporating 'being near the end of life' in the (risk-adjusted) compensation would reduce the predictable losses. One possibility to do so is to refine existing morbidity-based risk adjusters and/or to add new risk adjusters. In this respect, explicitly accounting for the multiple year character of the aftermath of a disease could prove helpful. For instance, individuals who had a stroke in a certain year do not only have above-average spending in that and the next year, but probably also in the years thereafter. This high spending in later years is unlikely to be fully captured by the risk-equalization model because morbidity flags tend to be based on utilization in the previous year. Another option is to use a form of cost-based compensation, for example the form applied in this paper (i.e. an ex-post compensation of actual spending by people who – in retrospect – turned out to be near end-of-life) or some form of outlier risk sharing (i.e. a compensation for X% of actual individual-level spending above a predefined threshold). Application of such cost-based compensations, however, come at a cost: although they can mitigate predictable profits and losses, they also reduce insurers' incentives for cost control. To guide the choice of policy measure(s) to better compensate for end-of-life spending, we recommend conducting a more in-depth analysis of specific health care of people near the end of life (which was not possible with the data available for this study). This might provide valuable insights in the type of ex-ante information available for better identification of those near end of life.

We did not find a statistically significant contribution of end-of-life spending to the predictable profits and losses for all subgroups of chronically ill. For some groups, like heart attack, this might be considered surprising. One explanation for this is that carving out end-of-life spending can have a direct and an indirect effect, which for some groups might balance each other out. The direct effect is that carving out end-of-life spending lowers overall spending within subgroups, reducing profits/losses. The indirect effect is that the payment weights of the risk adjusters in the risk-equalization model change. More specifically, carving out end-of-life spending lowers the payment weights for especially the morbidity-based risk adjusters, resulting in lower risk-adjusted payments. Another possible explanation for the absence of a statistically significant contribution of end-of-life spending for some groups, like diabetes, is that most of the chronically ill individuals in these groups are also flagged as such by the risk-equalization model (and the model also adequately captures the aftermath of a disease).

Our study comes with at least three limitations. Firstly, our findings based on the health survey data are conditional on the adult population of 19 years and older.

Moreover, the survey data are from a sample of the population and might suffer from some observation bias. We improved the representativeness of the survey as much as possible by applying a raking procedure. Nonetheless, some observation bias might remain. The third limitation regards the generalizability of our findings beyond the context of the Dutch health insurance system. The international relevance of our study can be found in the conclusion that – even under sophisticated risk equalization – end-of-life spending can significantly contribute to predictable profits and losses for selective groups. The groups for which this is true, however, will vary across healthcare systems and depend on contextual factors such as the benefits package, features of the risk equalization model, and characteristics of the relevant population. For example, the risk-equalization model used in Medicare Advantage can also be considered sophisticated, but unlike the Dutch model relies on concurrent rather than prospective morbidity flags. Moreover, the two models use different types of risk adjusters. Furthermore, Medicare Advantage pertains to the population of 65 and older, whereas the Dutch basic health insurance covers the entire population. Therefore, the exact contribution of end-of-life spending to predictable profits and losses in Medicare Advantage and other systems might differ from that in the Dutch setting and remains an empirical question. The methodology used in this paper could help to answer this question for healthcare systems in other countries.

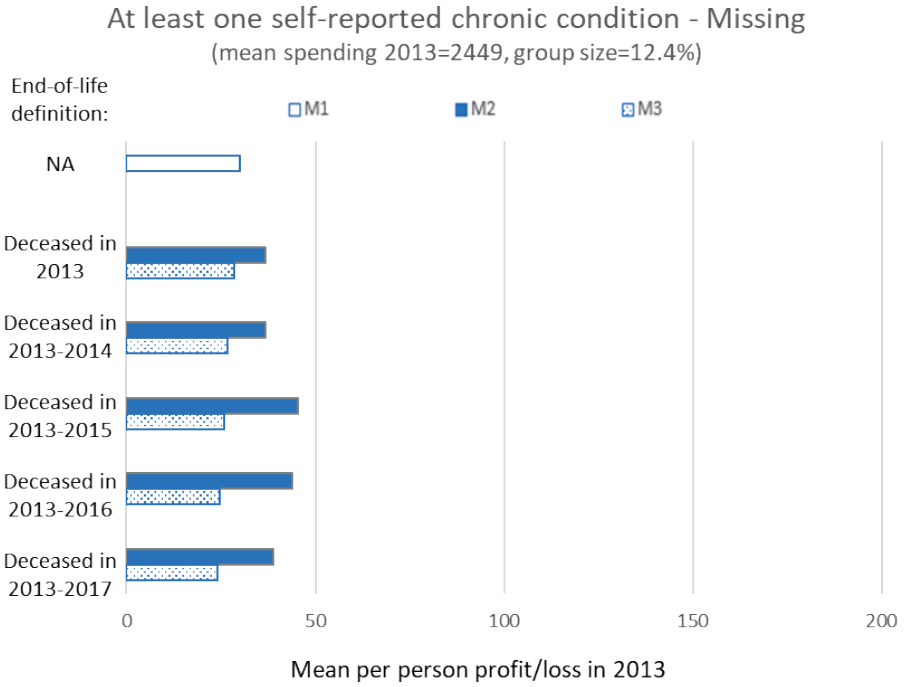
Regardless of whether end-of-life spending is accounted for, we find significant predictable profits and losses for all subgroups studied. This suggests that there are other factors that drive these profits and losses. To be able to further mitigate incentives for risk selection, it is important to identify these factors. Potential starting points could be analyses of the heterogeneity of the groups flagged by the morbidity-based risk adjusters in the risk-equalization model (i.e. how selective are these groups?), and of the reason(s) for why chronically ill individuals are not always identified as such by the risk-equalization model. Another option to further mitigate incentives for risk selection is using risk adjusters that are based on information from the current year instead of from a prior period. An example of a model using such concurrent risk adjusters can be found in the Marketplaces in the US. The advantage of a concurrent model over a prospective model as currently used in the Netherlands is a better model fit. However, a prospective model is better able to maintain and stimulate cost control incentives.

In this paper we assumed that risk selection will most likely take place on the basis of 'health' (e.g. yes/no chronic condition). An interesting question is whether risk selection might also take place based on yes/no being near end of life. For this to be

true, three preconditions must be met: 1) being near the end of life comes with a profit/loss for the insurer, 2) being near the end of life must be predictable to some extent and 3) insurers must be able to target this group with specific actions (Ellis & McGuire 2007; Van Kleef et al. 2019). Regarding the first condition, our results indicate a loss of approximately 5,700 euros for individuals who will die in the next four to five years, implying that this precondition is met. Regarding the predictability of being near end of life, previous research has clearly shown that it is hard to predict who will die in the upcoming year (Van Vliet & Lamers 1998; Einav et al. 2018). However, this might be different for those who will die in the next four or five years. Given the extensive medical information that has become increasingly available for specific conditions (e.g. cancer) and with the use of advanced statistical techniques (e.g. machine learning), it may well be possible to predict with reasonable certainty who will die within the coming five years. We believe this is an interesting and important direction for further research. The extent to which the third precondition is fulfilled, depends on the specific institutional characteristics of the health insurance market. In the Dutch context, insurers face open enrollment and a standardized benefits package, making it impossible for them to refuse specific individuals or to refrain from contracting certain health care. Although in theory insurers could engage in risk selection via the customer service and/or the contracted provider network, it is doubtful whether that would be effective given that individuals who are near the end of life are likely to have a low propensity to switch health plans.

APPENDIX A. Mean per person predictable profit/loss in 2013 under M1-M3 for five definitions of being near the end of life for the group for whom in 2012 the information on yes/no self-reported chronic condition is missing

Figure A.4.1 Mean per person profit in 2013 under M1-M3 for five definitions of being near the end of life for the group for whom in 2012 the information on yes/no self-reported chronic condition was missing.



Note: M1 is the Dutch risk-equalization model 2016, M2 is M1 supplemented with 100% cost-based compensation for people who are near the end of life, and M3 is M1 supplemented with a proportional cost-based compensation using the percentages in the right column of Table 4.2. NA means not applicable. An asterisk (*) next to a set of bars indicates a statistically significant ($p<0.05$) difference between the profit/loss under M2 versus M3.

APPENDIX B. Mean per person predictable loss in 2013 under three payment models, the absolute difference between M2 and M3 and the t-value for five definitions of being near the end of life for 19 chronic conditions individuals could report to have suffered from in the past 12 months

Table B.4.1 Mean per person loss in 2013 under M1-M3, the absolute difference between M2 and M3 and the t-value for five definitions of being near the end of life for 19 chronic conditions individuals could report in 2012 to have suffered from in the past 12 months.

Groups	End-of-life Definition	Mean per person profit/ loss under M1 (in euros)	Mean per person profit/ loss under M2 (in euros)	Mean per person profit/ loss under M3 (in euros)	Absolute difference between M2 and M3. * indicates statistically significant difference (p<0.05)	T-value of paired t-test between the mean per person profit/loss under M2 versus M3
Stroke (N=0.5% Mean spending=9093)	*NA*	-703	-	-	-	-
	Deceased in 2013	-	-781	-666	115	0.96
	Deceased in 2013-2014	-	-636	-629	7	0.05
	Deceased in 2013-2015	-	-535	-603	68	0.49
	Deceased 2013-2016	-	-474	-581	108	0.77
	Deceased in 2013-2017	-	-391	-562	172	1.21
Heart attack (N=0.4% Mean spending= 8504)	*NA*	-488	-	-	-	-
	Deceased in 2013	-	-637	-462	176	1.36
	Deceased in 2013-2014	-	-565	-437	128	0.85
	Deceased in 2013-2015	-	-544	-418	126	0.82
	Deceased 2013-2016	-	-599	-403	195	1.25
	Deceased in 2013-2017	-	-544	-390	154	0.84
Heart condition (N=2.1% Mean spending= 8919)	*NA*	-757	-	-	-	-
	Deceased in 2013	-	-659	-717	58	0.88
	Deceased in 2013-2014	-	-707	-678	29	0.42
	Deceased in 2013-2015	-	-627	-649	23	0.31
	Deceased 2013-2016	-	-746	-626	120	1.61
	Deceased in 2013-2017	-	-689	-605	84	1.11

Table B.4.1 Mean per person loss in 2013 under M1-M3, the absolute difference between M2 and M3 and the t-value for five definitions of being near the end of life for 19 chronic conditions individuals could report in 2012 to have suffered from in the past 12 months. (*continued*)

Groups	End-of-life Definition	Mean per person profit/ loss under M1 (in euros)	Mean per person profit/ loss under M2 (in euros)	Mean per person profit/ loss under M3 (in euros)	Absolute difference between M2 and M3. * indicates statistically significant difference ($p<0.05$)	T-value of paired t-test between the mean per person profit/loss under M2 versus M3
Cancer (N=1.7% Mean spending= 10474)	*NA*	-1130	-	-	-	-
	Deceased in 2013	-	-632	-1070	438*	3.89
	Deceased in 2013-2014	-	-378	-1012	634*	5.29
	Deceased in 2013-2015	-	-340	-970	630*	5.31
	Deceased 2013-2016	-	-191	-934	744*	5.93
	Deceased in 2013-2017	-	-192	-904	712*	5.78
Migraine (N=15.0% Mean spending= 2307)	*NA*	-107	-	-	-	-
	Deceased in 2013	-	-97	-101	4	0.30
	Deceased in 2013-2014	-	-95	-96	0.3	0.02
	Deceased in 2013-2015	-	-95	-92	3	0.23
	Deceased 2013-2016	-	-104	-88	15	1.05
	Deceased in 2013-2017	-	-98	-85	13	0.85
Blood pressure (N=16.0% Mean spending= 4432)	*NA*	-214	-	-	-	-
	Deceased in 2013	-	-214	-202	11	0.74
	Deceased in 2013-2014	-	-200	-191	9	0.52
	Deceased in 2013-2015	-	-194	-183	11	0.59
	Deceased 2013-2016	-	-219	-177	43*	2.32
	Deceased in 2013-2017	-	-222	-171	51*	2.79
Blood vessels (N=2.5% Mean spending= 7604)	*NA*	-611	-	-	-	-
	Deceased in 2013	-	-568	-578	10	0.19
	Deceased in 2013-2014	-	-502	-547	45	0.76
	Deceased in 2013-2015	-	-364	-524	160*	2.55
	Deceased 2013-2016	-	-360	-505	145*	2.31
	Deceased in 2013-2017	-	-325	-489	163*	2.54
Asthma (N=7.9% Mean spending=4732)	*NA*	-242	-	-	-	-
	Deceased in 2013	-	-185	-229	44	1.49
	Deceased in 2013-2014	-	-161	-217	56	1.72
	Deceased in 2013-2015	-	-183	-208	24	0.75
	Deceased 2013-2016	-	-180	-200	20	0.63
	Deceased in 2013-2017	-	-167	-194	26	0.80

Table B.4.1 Mean per person loss in 2013 under M1-M3, the absolute difference between M2 and M3 and the t-value for five definitions of being near the end of life for 19 chronic conditions individuals could report in 2012 to have suffered from in the past 12 months. (*continued*)

Groups	End-of-life Definition	Mean per person profit/ loss under M1 (in euros)	Mean per person profit/ loss under M2 (in euros)	Mean per person profit/ loss under M3 (in euros)	Absolute difference between M2 and M3. * indicates statistically significant difference (p<0.05)	T-value of paired t-test between the mean per person profit/loss under M2 versus M3
Psoriasis (N=2.7% Mean spending= 3780)	*NA*	-386	-	-	-	-
	Deceased in 2013	-	-366	-365	0.2	0.00
	Deceased in 2013-2014	-	-374	-345	28	0.80
	Deceased in 2013-2015	-	-357	-331	26	0.71
	Deceased 2013-2016	-	-376	-319	57	1.54
	Deceased in 2013-2017	-	-345	-309	36	0.93
Eczema (N=4.8% Mean spending= 2790)	*NA*	-151	-	-	-	-
	Deceased in 2013	-	-133	-143	10	0.45
	Deceased in 2013-2014	-	-122	-136	14	0.56
	Deceased in 2013-2015	-	-107	-130	23	0.87
	Deceased 2013-2016	-	-120	-125	5	0.20
	Deceased in 2013-2017	-	-70	-121	51	1.85
Severe/ recurring dizziness (N=4.1% Mean spending= 5819)	*NA*	-500	-	-	-	-
	Deceased in 2013	-	-484	-473	10	0.30
	Deceased in 2013-2014	-	-470	-447	22	0.58
	Deceased in 2013-2015	-	-432	-429	3	0.08
	Deceased 2013-2016	-	-465	-413	51	1.23
	Deceased in 2013-2017	-	-379	-400	20	0.47
Severe/ recurring disease of intestines (N=4.3% Mean spending= 5285)	*NA*	-575	-	-	-	-
	Deceased in 2013	-	-603	-545	58	1.36
	Deceased in 2013-2014	-	-534	-515	19	0.42
	Deceased in 2013-2015	-	-512	-493	19	0.40
	Deceased 2013-2016	-	-515	-475	40	0.87
	Deceased in 2013-2017	-	-501	-460	41	0.89
Incontinence (N=6.3% Mean spending= 5630)	*NA*	-238	-	-	-	-
	Deceased in 2013	-	-299	-225	74*	2.99
	Deceased in 2013-2014	-	-262	-213	49	1.65
	Deceased in 2013-2015	-	-228	-204	23	0.75
	Deceased 2013-2016	-	-251	-197	54	1.71
	Deceased in 2013-2017	-	-218	-190	28	0.84

Table B.4.1 Mean per person loss in 2013 under M1-M3, the absolute difference between M2 and M3 and the t-value for five definitions of being near the end of life for 19 chronic conditions individuals could report in 2012 to have suffered from in the past 12 months. (*continued*)

Groups	End-of-life Definition	Mean per person profit/ loss under M1 (in euros)	Mean per person profit/ loss under M2 (in euros)	Mean per person profit/ loss under M3 (in euros)	Absolute difference between M2 and M3. * indicates statistically significant difference ($p<0.05$)	T-value of paired t-test between the mean per person profit/loss under M2 versus M3
Wear of joint (N=13.2% Mean spending= 4870)	*NA*	-255	-	-	-	-
	Deceased in 2013	-	-262	-241	21	1.19
	Deceased in 2013-2014	-	-255	-228	27	1.38
	Deceased in 2013-2015	-	-227	-219	8	0.42
	Deceased 2013-2016	-	-239	-211	28	1.40
	Deceased in 2013-2017	-	-253	-204	49*	2.44
Joint inflammation (N=5.0% Mean spending= 5786)	*NA*	-334	-	-	-	-
	Deceased in 2013	-	-393	-317	76*	2.25
	Deceased in 2013-2014	-	-393	-299	94*	2.55
	Deceased in 2013-2015	-	-369	-287	82*	2.22
	Deceased 2013-2016	-	-376	-276	100*	2.69
	Deceased in 2013-2017	-	-373	-267	106*	2.82
Severe/ recurring condition of back (N=9.9% Mean spending= 4016)	*NA*	-232	-	-	-	-
	Deceased in 2013	-	-219	-219	0	0.01
	Deceased in 2013-2014	-	-207	-207	0	0.00
	Deceased in 2013-2015	-	-212	-199	14	0.57
	Deceased 2013-2016	-	-245	-191	53*	2.25
	Deceased in 2013-2017	-	-223	-185	38	1.56
Severe/ recurring condition of neck (N=9.4% Mean spending= 3731)	*NA*	-166	-	-	-	-
	Deceased in 2013	-	-137	-158	20	0.92
	Deceased in 2013-2014	-	-141	-149	8	0.37
	Deceased in 2013-2015	-	-145	-143	2	0.11
	Deceased 2013-2016	-	-171	-138	33	1.45
	Deceased in 2013-2017	-	-162	-133	29	1.26
Severe/ recurring condition of elbow (N=6.3% Mean spending= 4385)	*NA*	-167	-	-	-	-
	Deceased in 2013	-	-133	-159	25	0.85
	Deceased in 2013-2014	-	-161	-150	11	0.36
	Deceased in 2013-2015	-	-139	-144	5	0.14
	Deceased 2013-2016	-	-180	-138	41	1.29
	Deceased in 2013-2017	-	-177	-134	44	1.33

Table B.4.1 Mean per person loss in 2013 under M1-M3, the absolute difference between M2 and M3 and the t-value for five definitions of being near the end of life for 19 chronic conditions individuals could report in 2012 to have suffered from in the past 12 months. (*continued*)

Groups	End-of-life Definition	Mean per person profit/ loss under M1 (in euros)	Mean per person profit/ loss under M2 (in euros)	Mean per person profit/ loss under M3 (in euros)	Absolute difference between M2 and M3. * indicates statistically significant difference (p<0.05)	T-value of paired t-test between the mean per person profit/loss under M2 versus M3
Other (N=13.8% Mean spending =4889)	*NA*	-377	-	-	-	-
	Deceased in 2013	-	-337	-357	20	0.97
	Deceased in 2013-2014	-	-311	-338	27	1.15
	Deceased in 2013-2015	-	-301	-324	23	0.95
	Deceased 2013-2016	-	-301	-312	11	0.44
	Deceased in 2013-2017	-	-289	-302	13	0.50

Note: M1 is the Dutch risk-equalization model of 2016, M2 is M1 supplemented with 100% cost-based compensation for people who are near the end of life, and M3 is M1 supplemented with proportional cost-based compensation using the percentages in the right column of Table 4.2. Comparing the results for individuals who reported to have suffered a stroke in the past 12 months to the result for those who reported to have ever suffered a stroke (Figure 4.3), shows a statistically significant difference between M2 and M3 for the latter group, but not for the first. The reason for this can be found in a complex interplay of three factors: 1) the difference in mean spending of these groups in year t, 2) the difference in mean risk equalization payment of these groups in year t, and 3) the difference between these groups in terms of correlation with 'being near the end of life'. In particular the second factor might be relevant here since people who have suffered a stroke in the past 12 months are more likely to be flagged by a morbidity indicator (e.g., due to hospital treatment in year t-1) than those who have suffered a stroke ever in the past.

II

Using health survey information for the improvement of risk equalization


5

Incorporating self-reported health measures in risk equalization through constrained regression

With R.C. van Kleef & F. Eijkenaar

Based on: Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. (2020). Incorporating self-reported health measures in risk equalization through constrained regression. European Journal of Health Economics. https://doi.org/10.1007/s10198-019-01146-y

ABSTRACT

Most health insurance markets with premium-rate restrictions, include a risk equalization system to compensate insurers for predictable variation in spending. Recent research has shown, however, that even the most sophisticated risk equalization systems tend to undercompensate (overcompensate) groups of people with poor (good) self-reported health, confronting insurers with incentives for risk selection. Self-reported health measures are generally considered infeasible for use as an explicit 'risk adjuster' in risk equalization models. This study examines an alternative way to exploit this information, namely through 'constrained regression' (CR). To do so, we use administrative data (N=17m) and health survey information (N=380k) from the Netherlands. We estimate five CR models and compare these models with the actual Dutch risk equalization model of 2016 which was estimated by ordinary least squares (OLS). In the CR-models the estimated coefficients are restricted such that the under-/overcompensation for groups based on self-reported general health is reduced by 20, 40, 60, 80 or 100 percent. Our results show that CR can improve outcomes for groups that are *not* explicitly flagged by risk adjuster variables, but worsens outcomes for groups that *are* explicitly flagged by risk adjuster variables. Using a new standardized metric that summarizes under-/overcompensation for both types of groups, we find that the lighter constraints can lead to better outcomes than OLS.

5.1 INTRODUCTION

Many health insurance systems are based on the model of regulated competition. Competition among health insurers helps to improve efficiency of health insurance systems and regulation helps to protect public objectives like individual affordability of health plans. One element of the regulatory framework is risk equalization, a mechanism that compensates health insurers for predictable spending variation across individuals (Van de Ven & Ellis 2000; Van de Ven et al. 2007). In the presence of premium-rate restrictions, as applied in (almost) all regulated health insurance markets, risk equalization mitigates incentives for risk selection.

Over the past decades, risk equalization systems have evolved from simple demographic models to sophisticated health-based models. An example of the latter is the model applied in the Netherlands, which includes risk adjusters based on an extensive series of demographic, socioeconomic and morbidity-based variables. Even these sophisticated models, however, do not completely correct for predictable spending variation (Geruso & Layton 2015; McGuire et al. 2013; Van Kleef et al. 2017a). Van Kleef et al. (2019) find that the Dutch risk equalization model of 2016 undercompensates health insurers for the group of consumers who reported a fair or (very) poor health status in the prior year and overcompensates them for the group of consumers who reported a (very) good health status in the prior year. On average, the former group (about 24 percent of the population) confronts health insurers with a predictable loss of around 500 euros per person per year, while the latter (about 75 percent of the population) confronts them with a predictable profit of around 180 euros per person per year (Van Kleef et al. 2019).

Correlation between consumers' (self-reported) health and their profitability to health insurers can be problematic for the functioning of health insurance markets. When the unprofitable groups in poor health value (specific features of) health plans differently than the profitable groups in good health, health insurers are confronted with incentives to design their plans in a way that these are more attractive to healthy consumers than to unhealthy consumers. For instance, health insurers might refrain from contracting high-quality care for unprofitable groups with particular chronic medical conditions (Ellis & McGuire 2007; Ellis et al. 2017). These actions, which we refer to as *selection via plan design*, threaten the efficiency of health plans (Frank et al. 2000; Glazer & McGuire 2000; Han & Lavetti 2017; Pauly 1998; Schut & Varkevisser 2009; Van de Ven et al. 2015).

This paper seeks to mitigate incentives for selection via plan design by incorporating health survey information in the risk equalization model. However, *direct* use of self-reported health measures as a basis for risk adjusters is problematic, because the required survey information is not available for the entire population (which is typically considered a requirement for calculating individual-level risk equalization payments). Collecting this information for the entire population would usually be considered too cumbersome and costly (Van de Ven & Ellis 2000).

Although self-reported health measures are not appropriate as a basis for risk adjusters, they can be used *indirectly* in risk equalization models through the method of constrained regression (CR). Conventional risk equalization models are usually estimated by means of ordinary least squares (OLS). Given a set of risk adjusters, OLS results in coefficients that minimize the sum of squared residuals. CR allows for estimating coefficients that minimize the sum of squared residuals *conditional* on a pre-specified under- or overcompensation (for instance zero) for specific groups (Van Kleef et al. 2017b). Previous research has shown that application of CR can improve payment fit for groups not explicitly flagged by risk adjusters. At the same time, CR typically worsens payment fit for groups explicitly flagged by risk adjusters. Van Kleef et al. (2017b) have applied CR in the Dutch context for the risk equalization model 2015 and concluded that the improved payment fit for some groups can potentially outweigh the deteriorated payment fit for other groups.

The aim of this study is to examine and evaluate the use of health survey information in risk equalization through CR. To do so, we use administrative data and health survey information from the Netherlands. The administrative data are from 2013 and contain information on medical spending and risk adjuster variables for the entire Dutch population (N≈17m). These data are used to replicate the Dutch risk equalization model of 2016. Furthermore, we use health survey data from 2012 based on a large sample of the Dutch population (N≈387k). We estimate six models, that is, one base model estimated with OLS (i.e. the Dutch risk equalization model 2016) and five models estimated with CR.

Our empirical application comes with two methodological challenges. First, in order to meaningfully use health survey information as a basis for CR to improve risk equalization, this information must be representative for the population. As with most samples, this is not entirely the case for our survey sample. Prior studies have shown that this sample is somewhat healthier than the population (Van Kleef et al. 2017a; Volksgezonheidenzorg.info). We address this by rebalancing the sample using a raking procedure (Battaglia et al. 2009; Izrael et al. 2000) to correct for mismatches

between the sample and the population. Second, a metric is required to evaluate the outcomes of CR relative to OLS. We use a new standardized evaluation metric that summarizes under- and overcompensations for a cross tabulation of two types of groups, i.e. groups explicitly flagged by risk adjusters (for which previous research has demonstrated an *increase* in under-/overcompensation with CR compared to OLS) and groups not explicitly flagged by risk adjusters (for which previous research has shown a *decrease* in under-/overcompensation with CR compared to OLS). More specifically, we first calculate the total under-/overcompensation per group, take the absolute value of these total under-/overcompensations and then sum these over the relevant groups.

The structure of this paper is as follows. Section 5.2 describes relevant aspects of the Dutch health insurance system. Section 5.3 summarizes the relevant theory and previous research on selection via plan design and CR. Section 5.4 describes the data and methods for our empirical application and Section 5.5 presents the results. Finally, Section 5.6 summarizes and discusses the main findings.

5.2 THE DUTCH HEALTH INSURANCE MARKET

The Dutch health insurance market has two main components: a basic health insurance and a supplementary health insurance. Supplementary health insurance operates on the basis of free competition and is beyond the scope of this research. The basic health insurance operates on the basis of regulated competition. Regulations implemented by the Dutch government to ensure individual affordability and accessibility of the basic health insurance, include an individual mandate to buy basic health insurance, annual open enrollment, community-rated premiums, risk equalization and a standardized benefits package. The latter means that health plans have to cover a fixed set of benefits. Insurers are, however, free to selectively contract healthcare providers. Although this is intended to improve the efficiency of health care, health plans can also use this instrument to engage in selection via plan design, e.g. by not contracting good quality health care for specific unprofitable groups of consumers, also known as 'quality skimping' (Van Kleef et al. 2018a; Van de Ven et al. 2015).

Risk equalization mitigates incentives for selection via plan design, given premium-rate restrictions. The risk equalization model is used to calculate risk-adjusted payments to health plans, based on the characteristics of their insured population. The Dutch risk equalization model is comprised of three separate models: one for

somatic health care, one for mental health care and one model for copayments due to a mandatory deductible (Van Kleef et al. 2018a). This research focusses on the model for somatic health care, which contains the following indirect indicators of health: age, gender, region, socioeconomic status and source of income. In addition, the model includes the following series of more direct health indicators: pharmacy-based cost groups (PCGs), diagnosis-based cost groups (DCGs), multiple-year high cost groups (MHCGs), durable medical equipment cost groups (DMECGs), physiotherapy spending in the previous year, home care spending in the previous year and geriatric rehabilitation care spending in the previous year (Van Kleef et al. 2018a). In this paper, we refer to these direct health indicators as 'morbidity-based risk adjusters'.

5.3 LITERATURE REVIEW

5.3.1 Selection via plan design

The literature on (incentives for) selection via plan design originates from the work by Rothschild and Stiglitz (1976), who were the first to show theoretically that insurers react to adverse selection incentives and try to attract good risks through insurance plans' coverage and price. Glazer and McGuire (2000) applied this to the health insurance market and further developed the ideas of Rothschild and Stiglitz (1976) into a model of insurer and consumer behavior. Their model shows how profit-maximizing health insurers will engage in selection via plan design to attract good risks and deter bad ones, for example by creating networks in (dis)favor of some conditions and services. Breyer et al. (2012) called this 'indirect selection'. Furthermore, by applying the insights from Frank et al. (2000), Ellis and McGuire (2007) showed that health plan's incentives to engage in selection via plan design depend on both 'predictiveness' and 'predictability' (Ellis & McGuire 2007). Services have predictiveness if use of these services correlates with use of other services covered by the health plan. Services are predictable when consumers can (to some extent) predict how much of those services they will use during the contract period. When consumers take predicted use of services into account when choosing a health plan, they will be sensitive to differences in health plan design with regard to those services. Consequently, health insurers can influence the choice of consumers through health plan design (Ellis et al. 2013b; Han & Lavetti 2017; Layton et al. 2017; Lissenden 2019; McGuire et al. 2014). McGuire et al. (2014) added estimated demand elasticities to the predictiveness/predictability measures by studying incentives for selection via plan design in a market with risk adjustment, and again confirmed that health insurers have incentives to deter bad risks through health plan design,

specifically people with a chronic disease (McGuire et al. 2014). Ellis et al. (2017) concluded that incorporating demand elasticities across services is necessary to accurately assess incentives for selection via plan design.

Other studies have investigated the actual occurrence of selection via plan design in health insurance markets. For example, Cao and McGuire (2003) investigated the services offered by HMOs relative to the fee-for-service (FFS) sector within the Medicare program by researching the correlation between HMOs' market shares and the average expenditures in the FFS sector. Their hypothesis is as follows. If HMOs try to deter consumers who are more likely to use a service, i.e. high-risk individuals, they are expected to underprovide that service. Consequently, the HMOs will selectively enroll low-risk individuals with regard to that service. As more low-risk individuals enroll in HMOs, the average risk in the FFS sector will increase, resulting in higher average expenditures in the FFS sector. This all means that if service-level selection is present, the correlation between HMO market share and FFS average expenditures should be positive for services that the HMOs underprovide and negative for services they overprovide. Indeed, this is exactly what Cao and McGuire (2003) find. Also Eggleston and Bir (2009), Ellis et al. (2013b) and Decoralis and Gugliemo (2017) found evidence of health plans engaging in selection via plan design. Carey (2017a; 2017b), Lavetti and Simon (2019), Geruso et al. (2016) and Han and Lavetti (2017) found evidence for selection via plan design by health plans with regard to prescription drugs and Shepard (2016) showed the same for hospital network design.

In summary, existing literature suggests that incentives for selection via plan design are a function of consumers' expected spending for services covered by health plans and their (un)profitability to plans. Empirical studies have shown that insurers respond to these incentives via the design of their plans.

5.3.2 Constrained regression

Conventional risk adjustment models are typically estimated with OLS, which – given a set of risk adjusters – results in coefficients that minimize the residual sum of squares. CR allows for estimating coefficients that minimize the residual sum of squares *conditional* on a constraint imposed by the researcher. An example of a constraint is that the under- or overcompensation for a certain group equals a specific amount, such as zero (Van Kleef et al. 2017b).

Previous research has shown that – compared to OLS – use of CR in risk equalization comes with a trade-off between improved compensation for groups not explicitly flagged by risk adjusters and worsened compensation for groups explicitly flagged

by risk adjusters. In order to make a well-informed trade-off, Van Kleef et al. (2017b) argue that it is important to carefully define the groups that are vulnerable to risk selection. In addition, they argue that the relative importance of under-/overcompensations might vary with the size and sign (positive or negative) of the compensation. The authors find that under certain circumstances the improvement in compensation for groups not explicitly flagged by risk adjusters can outweigh the deterioration in compensation of groups that are explicitly flagged by risk adjusters (Van Kleef et al. 2017b).

Van Kleef et al. (2017b) were not the first to study the use of CR in the context of risk equalization. Glazer & McGuire (2002) already proposed including constraints in risk equalization to improve incentives for health plans. Starting from a model of insurer and consumer behavior, they showed that the optimal risk equalization coefficients result from CR with constraints for each of the separate services that health plans are able to distort. Layton et al. (2018) have empirically implemented this approach. A key difference between the present study and the studies mentioned above is that here the information used as a basis for constraints does not come from administrative data that is available for the entire population, but from a health survey that is only available for a sample of the population.

5.4 DATA AND METHODS

5.4.1 Data

To study the effects of including health survey information in the Dutch risk equalization model through CR, we merge administrative data from 2013 with health survey data from 2012. The administrative data comes from various administrative sources and contain information on individual-level medical spending and risk adjusters for all Dutch citizens with a basic health insurance in 2013 (n=16.9 million). The health survey data contain information on self-reported general health as well as specific self-reported chronic conditions for 387,195 individuals who were 19 years or older on September 1, 2012[12] and come from Statistics Netherlands (Volksgezonheidenzorg.info). The datasets were merged using a unique anonymized individual-level identification key.

12 Public Health Monitor (2012) of the Community Health Services, Statistics Netherlands and the National Institute for Public Health and the Environment.

5.4.1.1 Rebalancing

The survey sample used for this study is somewhat overrepresented by relatively healthy individuals (Van Kleef et al. 2017a; Withagen-Koster et al. 2018). To correct for differences in health as well as in age and socio-economic factors between the survey sample and the population, we rebalanced the sample by means of a raking procedure which was originally developed by Deming (1943). This procedure generates individual-level weights that equalize the frequencies of key variables in the sample to those in the population (Battaglia et al. 2009; Izrael et al. 2000). To see how this procedure works, imagine a sample that needs to be made representative for a population with respect to age and gender. As the joint distribution of these variables is only known in the sample, first a cross-tabulation of age and gender is made for the sample (say 20 categories for age and 2 for gender which results in 20 x 2= 40 cells). Next, for each separate row (say an age category), each entry of that row is multiplied by the ratio of the population total to the sample total for that age category, such that the row totals for the sample equal those for the population. So for each of the 20 rows/age categories, the 2 entries of gender are multiplied by the relevant ratio of the population total to the sample total. Then, this step is repeated for the columns (gender), after which the column totals will equal those in the population. The row totals (age categories), however, will no longer agree, although they are closer to the population totals than before the first iteration for the rows. This process is continued until agreement for both rows and columns is achieved (Battaglia et al. 2009; Izrael et al. 2000).

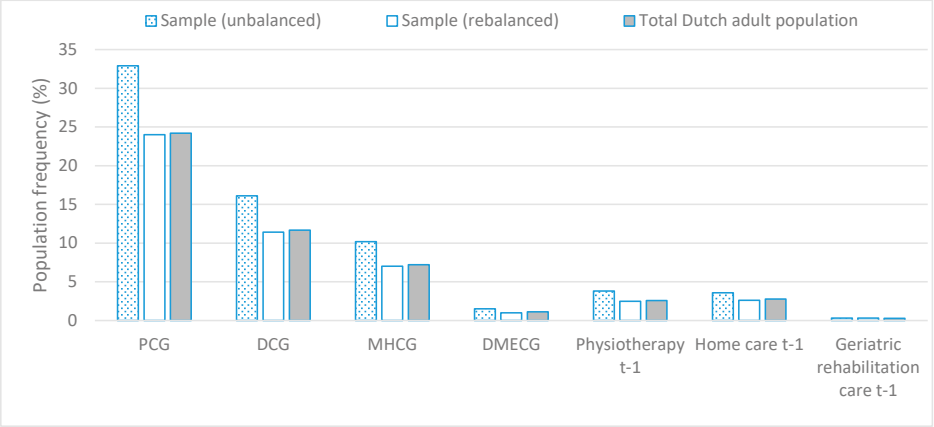
In addition to age and gender, our raking procedure includes all other risk adjuster classes of the risk equalization model 2016 (see Van Kleef et al. (2018a) for a complete list). Furthermore, the procedure also includes a proxy for whether or not someone had died in 2013 as well as 18 quantiles of mean total medical spending. The next section presents the representativeness of the sample before and after rebalancing.

5.4.1.2 Representativeness of survey sample

The survey sample includes 387,195 respondents of which 384,004 successfully merged with the administrative data of 2013. Unsuccessful matches can occur due to migration and death. After removing records with missing values on self-reported general health, 379,054 individuals remained for the analyses. Figure 5.1 shows the relative frequency in the sample and the total population, respectively, for the seven morbidity-based risk adjusters included in the risk equalization model 2016. Before rebalancing, the sample is overrepresented by people with morbidity. After rebalancing, the relative frequencies in the sample are close to those in the total population. For the same set of risk adjusters, Figure 5.2 shows the average spending

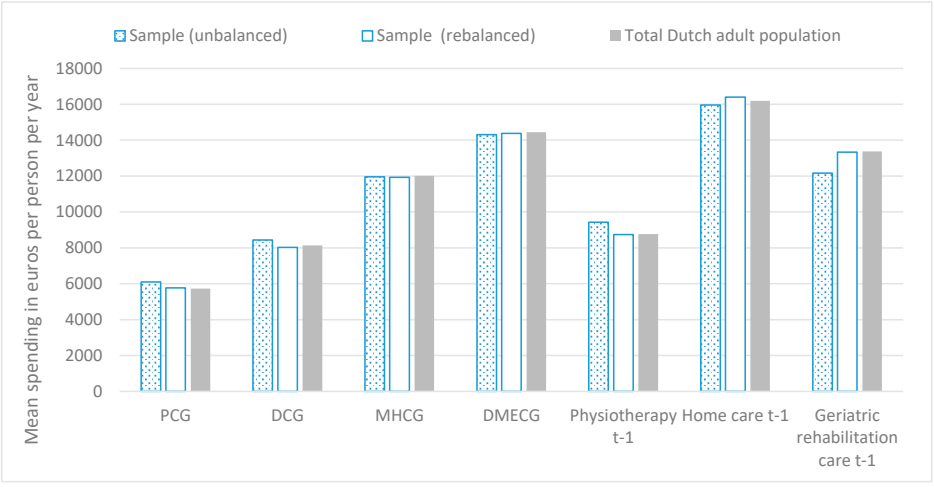
in the sample and the total population, respectively. Here too, the sample matches the population relatively well, especially after rebalancing. Appendix A shows similar patterns for other partitions of the population.

Figure 5.1. Frequencies per morbidity-based risk adjuster for the (un)balanced sample and total Dutch adult population



Note: The horizontal axis shows groups of individuals flagged by a morbidity-based risk adjuster included in the Dutch risk equalization model 2016. The abbreviations refer to pharmacy-based cost groups (PCGs), diagnosis-based cost groups (DCGs), multiple-year high cost groups (MHCGs) and durable medical equipment cost groups (DMECG).

Figure 5.2. Average spending per morbidity-based risk adjuster for the (un)balanced sample and total Dutch adult population



Note: The horizontal axis shows groups of individuals flagged by a morbidity-based risk adjuster included in the Dutch risk equalization model 2016. The abbreviations refer to pharmacy-based cost groups (PCGs), diagnosis-based cost groups (DCGs), multiple-year high cost groups (MHCGs) and durable medical equipment cost groups (DMECG).

5.4.1.3 Recalibrating the balanced survey sample

The Dutch risk equalization model is estimated with OLS. A property of OLS is that the mean predicted spending in the estimation dataset equals the mean spending in that dataset, implying that the residual spending has a mean of zero. This is not necessarily the case for subsamples drawn from the estimation dataset, such as our health survey sample. Before rebalancing, the average spending in the survey sample equals 3,190 euros and the predicted spending 3,223 euros, leaving a mean residual of -33 euros per person per year. After rebalancing, the mean spending equals 2,429 euros and the mean predicted spending equals 2,438 euros, leaving a mean residual of -9 euros. To correct for this remaining discrepancy, we recalibrated spending in the rebalanced sample by multiplying individual-level spending by a factor of 1.003767 (=2,438/2,429) so that the mean residual spending in the sample equals zero.

5.4.2 Methods

5.4.2.1 Model specification

To analyze the effects of incorporating self-reported health measures in risk equalization via CR, we estimated six models. The first model is the actual Dutch risk equalization model of 2016 (see Section 5.2) estimated with OLS. The other five models mimic the risk equalization model of 2016, but are estimated by CR. In the CR-models the under-/overcompensations of the group with a fair or (very) poor self-reported general health and the group with a (very) good self-reported general health are reduced by 20, 40, 60, 80 and 100 percent, respectively. Technically, imposing the constraints means that for each of the two groups of self-reported general health the sumproduct of risk adjuster values and estimated coefficients (i.e. total predicted spending for a group) equals a pre-specified amount. In an initial pass, the survey data are used to determine the risk adjuster values as well as the 'pre-specified amounts' corresponding to the abovementioned 20-40-60-80-100 percent reductions in under-/overcompensations for the relevant groups.

5.4.2.2 Evaluation

Prior research has shown that CR can improve compensation for some groups and worsen compensation for others. To evaluate the outcomes, we first calculated the under-/overcompensations based on the OLS model and the five CR-models for selected survey groups that are *not* explicitly flagged by risk adjusters in the risk equalization model, as well as for selected groups that *are* explicitly flagged by risk adjusters in that model. Under-/overcompensation is defined as the spending predicted by the risk equalization model minus the actual spending. The survey groups are based on self-reported general health, the number of self-reported chronic con-

ditions and specific self-reported chronic conditions (see Appendix B). Regarding the groups explicitly flagged by risk adjusters, we focus on those flagged by the seven morbidity-based risk adjusters (see Section 5.2).

Secondly, we constructed a standardized metric to evaluate the outcomes of CR compared to OLS in terms of group-level payment fit. Four groups are used for this part of the evaluation, i.e. yes/no (very) good self-reported general health (based on the health survey) cross tabulated with yes/no morbidity. The morbidity-group is defined as being flagged by at least one of the seven morbidity-based risk adjusters of the risk equalization model and the non-morbidity group is defined as being flagged by none of the seven morbidity-based risk adjusters. In this metric, we first calculate the total under-/overcompensation for the relevant groups. Next, we take the absolute values of these total under-/overcompensations and then sum these over the groups. For simplicity and interpretation purposes, we standardize the metric by taking the ratio of the outcome for a CR model to the outcome of the OLS model. Our measure can be written as follows:

$$S = \frac{\sum_{g=1}^{4}(\left|\sum_{i\in g} r_{cr,i}\right|)}{\sum_{g=1}^{4}(\left|\sum_{i\in g} r_{ols,i}\right|)} \quad (1)$$

Where:
iєg = the individuals belonging to group g;
$r_{cr,i}$ = the under-/overcompensation based on constrained regression for individual i;
$r_{ols,i}$ = the under-/overcompensation based on OLS for individual i;

When S>1, the outcomes of OLS are preferred over the outcomes of CR, while the opposite holds when S<1.

5.5 RESULTS

Section 5.5.1 presents and compares the outcomes of the six models in terms of individual-level fit. Section 5.5.2 presents the results under all six models for groups defined by self-reported general health and specific self-reported chronic conditions. The results for the groups explicitly flagged by the morbidity-based risk adjusters in the risk equalization model 2016 are presented in Section 5.5.3. Then, Section 5.5.4 presents the outcomes of the six models in terms of metric (1). In our primary analyses groups are weighted equally. Acknowledging that regulators might have reason to give more weight to some groups than to others, Section 5.5.5 illustrates the effects of a form of differentiated weighting.

5.5.1 Individual-level fit

Table 5.1 shows the individual-level *R*-squared[13] and Cummings' Prediction Measure (CPM)[14] for each model. As can be seen, for all CR models the R-squared is lower compared to that of the OLS model. Furthermore, the R-squared decreases as the constraint gets heavier. Under OLS, the residual sum of squares is minimized given the set of risk adjusters implying that – compared to OLS – any constraint of this type will result in a larger residual sum of squares. However, from the figures in Table 5.1 it can be concluded that imposing the constraints results in a very small reduction in payment fit at the individual level for both the R-squared and the CPM.

Table 5.1. Description and outcomes of the six models

Model	R-squared	CPM
OLS (0%)	27.9%	29.9%
CR-20%: Constrained regression model with 20 % reduction of under-/overcompensations for the two groups based on self-reported general health	27.9%	30.0%
CR-40%: Constrained regression model with 40 % reduction of under-/overcompensations for the two groups based on self-reported general health	27.8%	30.0%
CR-60%: Constrained regression model with 60 % reduction of under-/overcompensations for the two groups based on self-reported general health	27.8%	29.9%
CR-80%: Constrained regression model with 80 % reduction of under-/overcompensations for the two groups based on self-reported general health	27.7%	29.7%
CR-100%: Constrained regression model with 100 % reduction of under-/overcompensations for the two groups based on self-reported general health	27.6%	29.3%

5.5.2 Mean under-/overcompensation for groups identified in the survey data

To illustrate the effect of imposing the constraints, Table 5.2 presents the mean per person under-/overcompensation based on all six models for selected survey groups. As expected, under the CR models the under-/overcompensation for the two groups defined by self-reported general health (i.e. the groups on which the constraints are based) are reduced by 20, 40, 60, 80 and 100 percent compared to the OLS model.[15]

The per person undercompensation for the group with at least one self-reported chronic condition in the past year changes from -122 under OLS to -92 euros under

13 $\text{R-squared} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$

14 $\text{CPM} = 1 - \frac{\sum_{i=1}^{n}|y_i - \hat{y}_i|}{\sum_{i=1}^{n}|y_i - \bar{y}|}$

15 The under-/overcompensations do not exactly equal zero under the CR-100% model due to the recalibration of actual spending in the survey data (see Section 5.4.1.3).

CR-20% to 28 euros under CR-100%. A similar pattern can be observed for most of the other groups of chronically ill individuals. The group who has ever suffered from diabetes is on average even overcompensated by 538 euros per person per year under the CR-100% model, while OLS yields an undercompensation of 192 euros for this group. For the complementary groups of healthy individuals (i.e. those without the respective chronic condition(s)), the overcompensation generated by OLS mostly changes to an undercompensation under the CR-100% model. For example, for the group who reported no chronic condition in the last 12 months, the overcompensation of 178 euros under OLS turns into an undercompensation of 60 euros under the CR-100% model. Table 5.2 shows that the under- and overcompensations for all groups change linearly across the different CR models.

Table 5.2 also reports results for groups of survey respondents for whom the relevant information is missing. As can be seen, these groups have higher mean spending than the groups without the relevant chronic condition(s). In addition, the change in compensation when moving from OLS to CR follows the same pattern as that of the chronically ill groups, indicating that the missing groups are overrepresented by relatively unhealthy individuals.

Appendix B shows the same results for the 19 specific chronic conditions that survey respondents reported (not) to be suffering from in the past 12 months. Again, the chronically ill groups receive more compensation under CR than under OLS, while the overcompensations for the healthy counterparts decrease slightly.

5.5.3 Mean under-/overcompensation for groups flagged by morbidity-based risk adjusters in the risk equalization model

Table 5.3 presents the mean per person under-/overcompensation under all six models for yes/no morbidity as well as separately for the seven morbidity-based risk adjusters in the risk equalization model of 2016. The mean under-/overcompensation for all groups is zero under OLS, except for the PCG group. The reason for this is that the PCG classes are not mutually exclusive, while the classes within all other risk adjusters are. For all other morbidity-based risk adjusters, the mean compensation under OLS is zero as this is a property of OLS. Under CR, however, this is no longer the case. As Table 5.3 shows, the compensation for the groups with morbidity increases as the constraint becomes heavier. Under the CR-100% model, all groups explicitly flagged by a morbidity-based risk adjuster have a mean overcompensation of at least 600 euros and the entire group of individuals flagged by at least one morbidity-based risk adjuster (yes morbidity) has a mean overcompensation of 548 euros. In contrast, the compensation for all the complementary groups of healthy

Table 5.2. Mean under-/overcompensation by six models in euros per person per year for groups identified in the health survey

Survey group		Size of group	Mean spending in euros (2013)	Mean under-/overcompensation per person per year in euros (2013)					
				OLS (0%)	CR-20%	CR-40%	CR- 60%	CR-80%	CR-100%
Self-reported general health	Fair, poor or very poor	27.6%	5602	-494*	-396*	-297*	-198*	-100*	-1
	Good or very good	72.4%	1439	156*	125*	93*	62*	31*	-1
Self-reported chronic condition (past 12 months)	At least one	60.1%	3376	-122*	-92*	-62*	-32*	-2	28*
	None	28.2%	1010	178*	130*	83*	35*	-13*	-60*
	1	25.6%	2182	50*	41*	31*	22*	13	4
	2	15.1%	3095	-126*	-105*	-83*	-61*	-39*	-17
	3	8.4%	4352	-348*	-289*	-229*	-169*	-109*	-49*
	4	11.0%	6443	-427*	-297*	-166*	-36*	95*	226*
	Missing	11.7%	2396	26*	29*	33*	37*	41*	45*
Diabetes (ever)	Yes	8.2%	6739	-192*	-46*	99*	246*	392*	538*
	No	86.9%	2116	16	4	-7	-18	-28*	-39*
	Missing	5.0%	3077	-47*	-27*	-6	14	34*	55*
Stroke (ever)	Yes	4.2%	7626	-811*	-686*	-561*	-435*	-310*	-184*
	No	91.6%	2251	28*	23*	18	14	9	4
	Missing	4.2%	2903	-57*	-42*	-27*	-12	3	18
Heart attack (ever)	Yes	5.0%	7631	-456*	-320*	-184*	-47*	89*	225*
	No	90.9%	2241	19	13	8	3	-3	-8
	Missing	4.1%	2955	-76*	-61*	-45*	-29*	-14	2
Cancer (ever)	Yes	9.9%	6517	-433*	-351*	-270*	-188*	-106*	-24
	No	86.3%	2122	34*	26*	20*	13	6	-1
	Missing	3.8%	2739	-33*	-21	-10	2	14	26*

*= statistically significantly different from zero (P<0.05).
Note: The abbreviations stand for ordinary least squares (OLS) and constrained regression (CR)

people decreases. Under the CR-100% model, the entire group of individuals not flagged by a morbidity-based risk adjuster is on average undercompensated by 183 euros per person per year.

Table 5.3. Mean under-/overcompensation by six models in euros per person per year for groups (not) flagged by the morbidity-based risk adjusters of the risk equalization model 2016

Group		Size of group	Mean spending in euros (2013)	Mean under-/overcompensation per person per year in euros (2013)					
				OLS (0%)	CR-20%	CR-40%	CR- 60%	CR-80%	CR-100%
Morbidity	Yes	25.0%	5584	2	111*	220*	330*	439*	548*
	No	75.0%	978	-1	-37*	-74*	-110*	-147*	-183*
PCG	Yes	19.3%	5669	15*	134*	255*	375*	496*	616*
	No	80.7%	1286	-3*	-32*	-61*	-90*	-118*	-147*
DCG	Yes	9.3%	8179	0	145*	291*	437*	583*	729*
	No	90.7%	1514	0	-15*	-30*	-45*	-59*	-74*
MYHCG	Yes	5.8%	12137	0	211*	423*	634*	846*	1057*
	No	94.2%	1524	0	-13*	-26*	-38*	-51*	-64*
DMECG	Yes	0.9%	14727	0	167*	335*	502*	670*	838*
	No	99.1%	2020	0	-1	-3	-4*	-6*	-7*
Physiotherapy t-1	Yes	2.0%	8769	0	156*	313*	470*	627*	784*
	No	98.0%	1998	0	-3*	-6*	-9*	-13*	-16*
Home care t-1	Yes	2.2%	16658	0	231*	463*	695*	927*	1158*
	No	97.8%	1827	0	-5*	-10*	-15*	-19*	-24*
Geriatric rehabilitation care t-1	Yes	0.2%	13372	0	210*	422*	633*	844*	1055*
	No	99.8%	2109	0	0	-1	-1	-2	-2

*= statistically significantly different from zero (P<0.05).

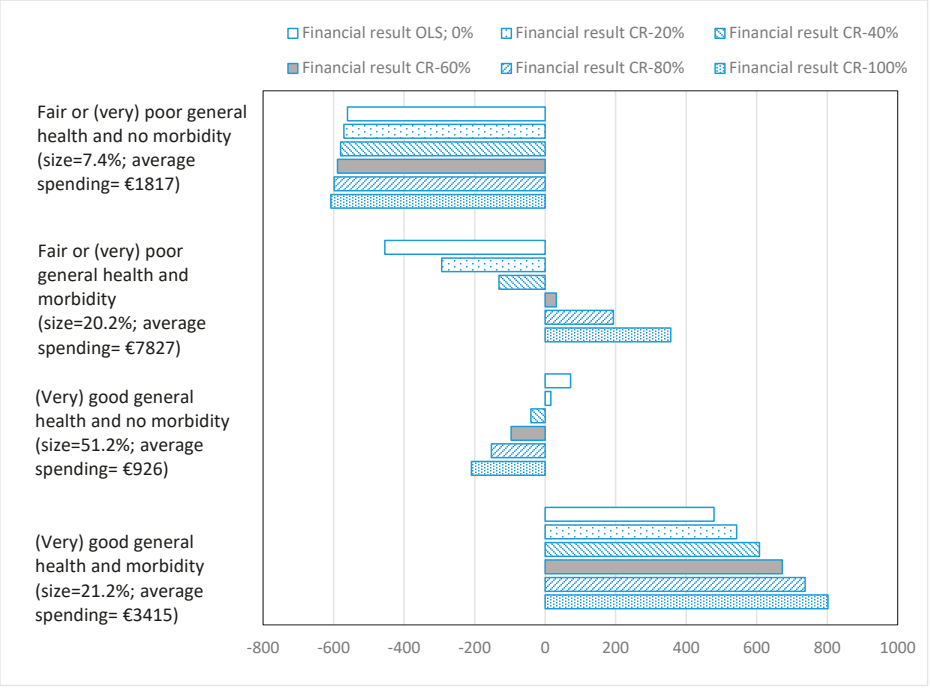
Note: The abbreviations stand for ordinary least squares (OLS), constrained regression (CR), pharmacy-based cost groups (PCGs), diagnosis-based cost groups (DCGs), multiple-year high cost groups (MHCGs) and durable medical equipment cost groups (DMECG). Morbidity is defined as being classified in one of the seven morbidity-based risk adjusters of the risk equalization model. No morbidity is defined as being classified in none of the seven morbidity-based risk adjusters of the risk equalization model 2016

5.5.4 Mean per person under-/overcompensation for a cross tabulation of groups identified in the survey data and groups flagged by morbidity-based risk adjusters in the risk equalization model

Even though average compensation increases for chronically ill groups, this is not necessarily the case for subsamples of these groups. To illustrate this, Figure 5.3 cross tabulates the two groups based on self-reported general health and the two groups based on yes/no morbidity as identified by the risk equalization model. The results show that within the group with a fair or (very) poor self-reported general

health, the individuals flagged by a morbidity indicator (20.2 percent of population) receive more compensation under CR than under OLS (i.e. from -455 euros under the OLS (0%) to 356 euros per person per year under the CR-100% model). However, the group of individuals who reported their health to be fair or (very) poor but who are not flagged by a morbidity indicator (7.4 percent of population), receive slightly less compensation under CR. As a result, the compensation for this subgroup decreases from -564 euros under OLS (0%) to -608 euros per person per year under the CR-100% model. A similar pattern can be observed within the group of people who reported a (very) good general health. The individuals who reported a (very) good general health and who are flagged by a morbidity indicator in the risk equalization model (21.2 percent of population), receive more compensation under CR than under OLS (i.e. from 479 euros under OLS (0%) to 802 euros per person per year under the CR-100% model). The individuals who reported a (very) good general health but are not flagged by a morbidity indicator (51.2 percent of population) receive less compensation under CR than under OLS (i.e. from 72 euros under OLS (0%) to -209 euros per person per year under the CR-100% model).

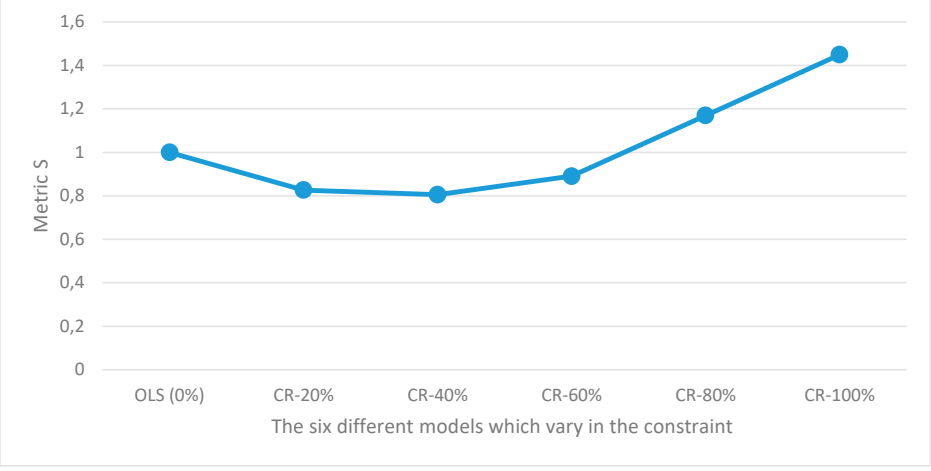
Figure 5.3. Mean under-/overcompensations under six models in euros per person per year for groups based on a cross tabulation of self-reported general health by yes/no morbidity



Note: The abbreviations stand for ordinary least squares (OLS) and constrained regression (CR). Morbidity is defined as being classified in one of the seven morbidity-based risk adjusters of the risk equalization model. No morbidity is defined as being classified in none of the seven morbidity-based risk adjusters of the risk equalization model.

In Figure 5.3 it is not obvious which of the models leads to the best outcomes overall. Figure 5.4 presents the outcomes of metric S (equation 1) which summarizes the outcomes over the four subgroups presented in Figure 5.3. This metric first calculates the total under-/overcompensation per group, takes the absolute value of these total under-/overcompensations and sums these over the four groups. The metric compares the outcomes of a CR-model relative to OLS. When S<1 a CR-model outperforms OLS, while S>1 implies the opposite. Figure 5.4 shows that S<1 for the CR-20%, CR-40% and CR-60% models, indicating that these models perform better than OLS with respect to the groups analyzed here. The CR-40% model has the lowest S-value (i.e. 0.81), indicating that overall this model performs best (given our choice of groups). For the CR-80% and CR-100% models S>1, indicating that these models perform worse than OLS (0%), with respect to these groups. In addition, we also see that CR in risk equalization can be pushed too far: applying a stricter constraint can cause the S-value to increase again.

Figure 5.4. Outcomes of six models for metric S for groups based on a cross tabulation of self-reported general health and yes/no morbidity



Note: The abbreviations stand for ordinary least squares (OLS) and constrained regression (CR). Metric S is calculated using equation (1). The constraint is a % reduction in under- and overcompensation on the group with a (very) good general health and the group with a fair or (very) poor general health.

5.5.5 Differentiated weighting of subgroups

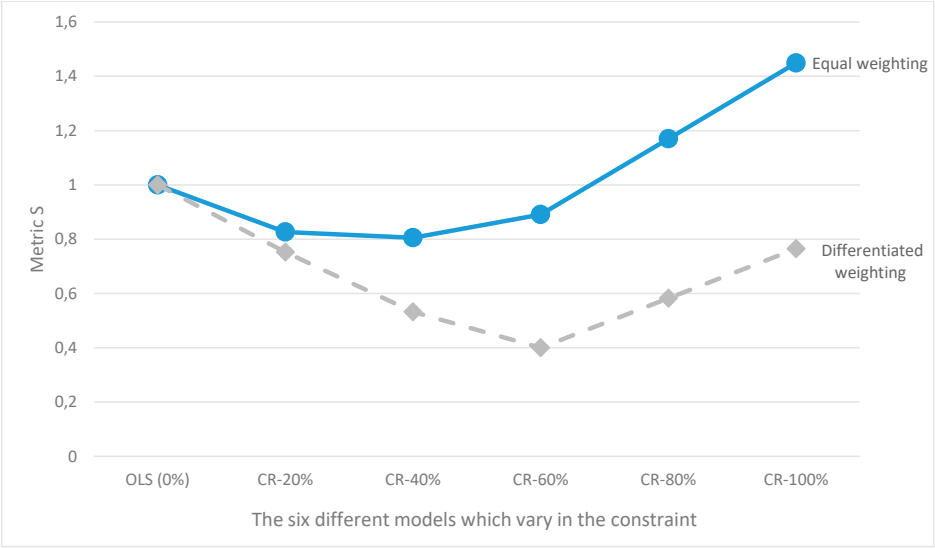
In Figure 5.4 (the under-/overcompensations of) all four groups are weighted equally. In practice, however, regulators might have reason to give more weight to some groups than to others. This might be the case when the regulator believes that some selection actions are more harmful than others. Section 5.3 supports this as it shows that under-/overcompensation and the resulting selection actions for some groups

might be more problematic compared to others. For example, the regulator might consider 'quality skimping' through selection via plan design to be more harmful for the functioning of the healthcare system than 'selective marketing'. In such a situation the regulator might give more weight to groups that are particularly vulnerable to 'quality skimping' (e.g. groups of chronically ill people with high expected spending) than to groups that are more likely to be subject to 'selective marketing' (e.g. groups of healthy people). In addition, the regulator might consider an undercompensation, which incentivizes insurers to underserve people, to be more harmful than an overcompensation, which incentivizes insurers to overserve people. Although it is not our goal here to advocate a specific form of differentiated weighting of subgroups, we believe it is instructive to indicate how weighting could influence the outcomes of the models simulated here.

Figure 5.5 compares the results of the CR-models under equal-weighting of subgroups with those under differentiated weighting of subgroups. The data series 'equal-weighting' is equivalent to the results of Figure 5.4. The data series 'differentiated weighting' presents the results of the CR-models relative to OLS with two types of differentiated weighting: (1) groups with high expected spending are given more weight than those with low expected spending (in our illustration: through weighting with the average spending of the groups) and (2) undercompensations are given more weight than overcompensations (in our illustration: through weighting an undercompensation twice as heavy as an overcompensation). A regulator might consider the first type of weighting when it is particularly concerned about quality skimping, for instance through selection via plan design. A regulator might think about the second type of weighting when 'underserving' is considered more harmful than 'overserving'.

Figure 5.5 shows that the line for 'differentiated weighting' of subgroups lies below the line of 'equal-weighting' of subgroups. This indicates that differentiated weighting of subgroups can substantially affect the outcomes of constrained regression compared to OLS. The results in Figure 5.5 also show that under 'differentiated weighting' the lowest value for S is to be found for the CR-60% model instead of the CR-40% model. This indicates that the optimal specification of a constraint can be affected by how a regulator weights the subgroups of interest.

Figure 5.5. Outcomes of six models for metric S with equal weighting and differentiated weighting of subgroups



Note: The outcomes with equal weighting of subgroups are calculated using equation (1). Differentiated weighting means that the result for each of the four groups is weighted with the average spending of that group and that an undercompensation is weighted twice as heavy as an overcompensation. The horizontal axis represents the different models. The series 'equal weighting' is equivalent to the outcomes of figure 5.4.

5.6 DISCUSSION

Most health insurance markets with premium-rate restrictions include a risk equalization system to compensate health insurers for predictable variation in spending. Recent research has shown, however, that even the most sophisticated risk equalization systems tend to undercompensate (overcompensate) people with poor (good) self-reported health, which confronts insurers with selection incentives. Self-reported health measures are generally considered infeasible for use as 'risk adjusters' in the risk equalization model. The aim of this paper was to examine and evaluate an alternative way of including self-reported health measures in risk equalization, namely through constrained regression (CR). To do so, we estimated five CR models and compared these with the actual Dutch risk equalization model of 2016 estimated with ordinary least squares (OLS). In the CR models, coefficients were estimated by least squares regression given that the under-/overcompensation for two groups based on self-reported general health are reduced by 20, 40, 60, 80 or 100 percent.

We first calculated the under- and overcompensations for selected survey groups and groups flagged by the morbidity-based risk adjusters included in the risk equal-

ization model. For the survey groups the results showed that the chronically ill receive more compensation under CR compared to OLS while the opposite is true for the complementary groups of healthy people. We observed a similar pattern for the groups (not) explicitly flagged by a morbidity-based risk adjuster; the groups that were explicitly flagged by such a risk adjuster receive more compensation under CR compared to OLS and the groups not explicitly flagged receive less. Next, we researched subsamples of these groups by cross tabulating the groups yes/no (very) good self-reported general health with the groups yes/no explicitly flagged by at least one morbidity-based risk adjuster. The results showed that – compared to OLS – also within the groups of self-reported general health the CR models move money from the individuals not flagged by a morbidity-based risk adjuster to those flagged by such a risk adjuster. Consequently, we found that payment fit improves for some groups but worsens for others. Van Kleef et al. (2018b) reported similar findings.

In order to evaluate the outcomes under all six models, we constructed a standardized metric that summarizes the absolute under-/overcompensations for relevant subgroups. We evaluated the four groups resulting from the cross tabulation of yes/no (very) good self-reported general health with the groups yes/no explicitly flagged by at least one morbidity-based risk adjuster. In this metric, we take the absolute values of the total under-/overcompensations and sum these over the four groups. The metric then compares the outcomes of a CR-model relative to OLS. We find that the CR-20%, CR-40% and CR-60% models yield more preferable outcomes than OLS, with the CR-40% model yielding the best results (i.e. for the groups analyzed here). This finding shows that a relatively small constraint could already improve conventional risk equalization. This is in line with the conclusions drawn in the paper by Van Kleef et al. (2017b) and with the findings of the work by Glazer & McGuire (2000; 2002). Glazer & McGuire (2000; 2002) argued that conventional risk equalization estimated with OLS might not be optimal and that overpaying groups flagged as 'high risk' and underpaying groups of 'low risk' could improve the outcomes of risk equalization. However, our results also show that CR in risk equalization can be pushed too far since the metric increases sharply as the constraint becomes heavier, with the CR-80% and CR-100% models performing worse than OLS.

Our primary simulations assume equal weighting of (the under-/overcompensations of) subgroups. Acknowledging that regulators might consider the effects of some selection actions to be more harmful than others, we also examined how differentiated weighting could influence the model outcomes. We found that a specific form of weighting (based on assumptions about the effects of quality skimping and underserving versus overserving) substantially affects the outcomes of the CR models

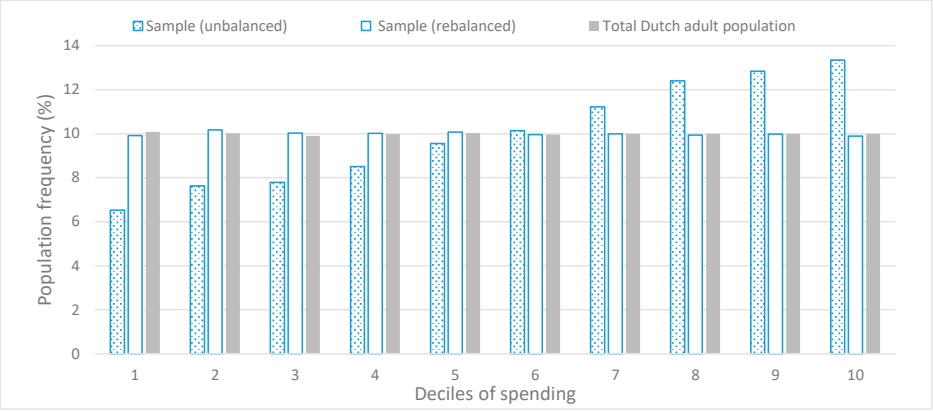
relative to OLS. These results demonstrate the relevance of carefully defining the policy objectives which regulators want to include in the evaluation.

The results of this study indicate that the use of health survey information in risk equalization through CR can be promising in reducing incentives for selection via plan design. Practical implementation of survey information in risk equalization through CR, however, needs more work. First, evaluation can be more refined, for example by evaluating the outcomes using other and more groups than analyzed here. In addition, more refined evaluation of risk equalization models could require a welfare approach that incorporates how incentives affect the behavior of insurers, how this behavior of insurers interacts with the behavior of consumers and how this affects social welfare. Although such a welfare-approach is beyond the scope of this paper, we believe that further research into this direction can help to improve the evaluation of risk equalization systems. Second, the choice of groups on which the constraints are based can differ from the groups used is this research. This choice is, however, not ours to make. The method of CR offers regulators an effective tool for protecting specific groups of interest against selection via plan design (Van Kleef et al. 2017b). An important insight in this respect is that these groups can also be determined on subsamples of the population, as long as these subsamples are representative for the population.

APPENDIX A: Representativeness of the survey sample by decile of spending

The representativeness of the sample (before and after rebalancing) for the population can also be shown by using deciles of spending instead of the morbidity adjusters. Figure A.5.1 shows the relative frequency per decile of actual medical spending for the unbalanced sample (dotted bars), rebalanced sample (empty bars) and the total adult population (solid bars). Before rebalancing, the sample is clearly overrepresented by high-spenders. After rebalancing, the relative frequency per decile of spending in the survey sample matches that of the population.

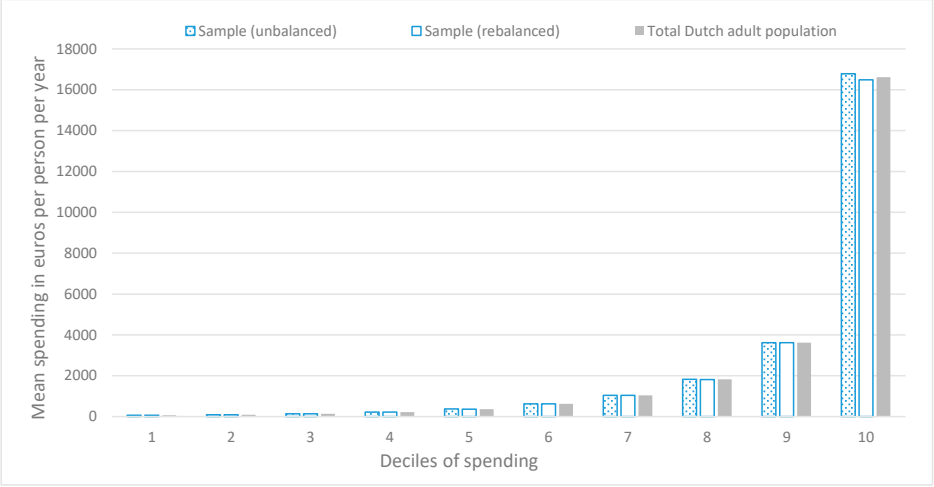
Figure A.5.1. Frequencies per decile of spending for the (un)balanced sample and the total Dutch adult population



Note: the deciles on the horizontal axis have been determined on the total Dutch adult population. The bars for the total Dutch adult population and the rebalanced sample do not exactly equal ten percent due to rounding.

Figure A.5.2 presents the average spending per decile of spending. Patterns in the sample are in line with those in the population, before but especially after rebalancing.

Figure A.5.2. Average spending per decile of spending for the (un)balanced sample and the total Dutch adult population



Note: the deciles on the horizontal axis have been determined on the total Dutch adult population.

APPENDIX B: Under-/overcompensation for survey groups regarding self-reported conditions under OLS and constrained regression

Table B.5.1. Under-/overcompensations in euros for survey groups of specific self-reported conditions in the last 12 months estimated with ordinary least squares (OLS) and constrained regression (CR)

Groups		Size	Mean spending in euros	Mean under-/overcompensation per person per year in euros (2013)					
				OLS (0%)	CR-20%	CR-40%	CR- 60%	CR-80%	CR-100%
Stroke	Yes	0,7%	9117	-722*	-561*	-399*	-237*	-75*	87*
	No	95,3%	2368	9	7	5	3	1	-1
	Missing	4,0%	3074	-92*	-74*	-57*	-39*	-21	-4
Heart attack	Yes	0,5%	8509	-503*	-333*	-162*	9	180*	351*
	No	95,3%	2376	7	4	3	1	-1	-3
	Missing	4,2%	3102	-82*	-62*	-43*	-23*	-4	16
Heart condition	Yes	3,2%	8918	-790*	-630*	-469*	-309*	-148*	12
	No	92,6%	2267	18	13	8	4	-1	-6
	Missing	4,2%	2908	-7	10	27*	44*	62*	79*
Cancer	Yes	2,8%	10444	-1137*	-1022*	-905*	-789*	-673*	-557*
	No	93,0%	2260	23*	20*	17	14	10	7
	Missing	4,2%	2969	-45*	-28*	-11	6	23*	40*
Migraine	Yes	12,4%	2286	-107*	-98*	-89*	-81*	-72*	-63*
	No	74,5%	2341	30*	24*	17	11	5	-1
	Missing	13,1%	3193	-48*	-23	2	28*	53*	79*
Blood pressure	Yes	22,0%	4415	-208*	-150*	-91*	-32*	27	86*
	No	65,4%	1897	49*	32*	15	-2	-19*	-35*
	Missing	12,7%	3035	-17	4	26*	48*	69*	91*
Blood vessels	Yes	3,6%	7507	-552*	-421*	-291*	-160*	-29	102*
	No	83,6%	2198	20*	13	5	-2	-9	-16
	Missing	12,7%	3081	-28*	-5	18	41*	64*	88*
Asthma	Yes	8,6%	4702	-263*	-182*	-100*	-19	63*	144*
	No	78,9%	2127	28*	16	5	-6	-18	-29*
	Missing	12,5%	3035	-16	6	28*	51*	73*	95*
Psoriasis	Yes	2,9%	3743	-383*	-349*	-314*	-280*	-245*	-211*
	No	83,8%	2290	13	7	2	-3	-8	-13
	Missing	13,3%	3156	-5	21	48*	74*	101*	127*
Eczema	Yes	4,2%	2758	-150*	-136*	-122*	-108*	-94*	-80*
	No	83,0%	2316	15	10	6	1	-3	-8
	Missing	12,8%	3142	-44*	-20	5	30*	54*	79*

Table B.5.1. Under-/overcompensations in euros for survey groups of specific self-reported conditions in the last 12 months estimated with ordinary least squares (OLS) and constrained regression (CR) (*continued*)

Groups		Size	Mean spending in euros	Mean under-/overcompensation per person per year in euros (2013)					
				OLS (0%)	CR-20%	CR-40%	CR- 60%	CR-80%	CR-100%
Severe/ recurring dizziness	Yes	4,3%	5772	-502*	-400*	-299*	-197*	-95*	7
	No	82,9%	2180	28*	19*	11	2	-6	-15
	Missing	12,8%	3095	-27*	-3	21	45*	68*	92*
Severe/ recurring disease of intestines	Yes	4,4%	5273	-590*	-514*	-438*	-362*	-285*	-209*
	No	83,1%	2202	34*	26*	19	12	4	-3
	Missing	12,4%	3076	-26*	-3	20	43*	66*	90*
Incontinence	Yes	8,2%	5579	-240*	-154*	-67*	19	105*	191*
	No	79,0%	2099	24*	13	3	-7	-17	-28*
	Missing	12,8%	3085	-36*	-13	10	34*	57*	80*
Wear of joint	Yes	18,6%	4858	-267*	-191*	-115*	-39*	37*	113*
	No	69,2%	1929	48*	31*	14	-2	-19*	-36*
	Missing	12,2%	2982	-10	11	32*	53*	74*	95*
Joint inflammation	Yes	6,4%	5804	-354*	-241*	-127*	-14	100*	214*
	No	80,8%	2143	24*	13	3	-8	-18	-28*
	Missing	12,8%	3076	-18	5	29*	52*	76*	100*
Severe/ recurring condition of back	Yes	11,0%	3987	-233*	-174*	-115*	-55*	4	64*
	No	76,5%	2151	33*	21*	10	-1	-12	-23*
	Missing	12,5%	3023	-20	2	25*	47*	69*	92*
Severe/ recurring condition of neck	Yes	10,1%	3706	-171*	-117*	-63*	-9	45*	100*
	No	77,3%	2191	25*	15	5	-5	-15	-25*
	Missing	12,5%	3064	-32*	-10	12	34*	56*	78*
Severe/ recurring condition of elbow	Yes	6,9%	4378	-184*	-112*	-39*	33*	106*	179*
	No	80,4%	2193	20*	10	1	-8	-17	-26*
	Missing	12,7%	3100	-39*	-17	7	30*	53*	76*
Other	Yes	15,0%	4860	-375*	-311*	-247*	-183*	-118*	-54*
	No	74,6%	1900	74*	58*	43*	27*	12	-3
	Missing	10,4%	3092	-34*	-10	14	38*	62*	86*

*= statistically significantly different from zero (P<0.05).

6

An innovative method to mitigate risk selection incentives in health insurance markets with sophisticated risk equalization: High-risk pooling based on health survey information

With R.C. van Kleef & F. Eijkenaar

Based on: Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. An innovative method to mitigated risk selection incentives in health insurance markets with sophisticated risk equalization: high-risk pooling based on health survey information. Submitted

ABSTRACT

Despite sophisticated risk equalization, insurers in regulated health insurance markets still face incentives to attract healthy people and avoid the chronically ill because of large differences in profitability between these groups. In principle, the best approach to mitigate such incentives for risk selection is to improve the risk-equalization model by adding or refining risk adjusters. However, not all potential risk adjusters are appropriate. One example is health survey information: despite its predictiveness of future healthcare spending, such information is generally considered inappropriate for risk equalization. We study high-risk pooling (HRP) as a promising alternative strategy to exploit health survey information for mitigating risk selection incentives. The essence of our HRP concept is that insurers can ex-ante assign predictably unprofitable individuals to a 'high risk pool' using information from a health survey. We evaluate the effect of five alternative pool sizes based on predicted residual spending post risk equalization on insurers' incentives for risk selection and cost control, and compare this to the situation without HRP. Our main conclusion is that HRP has the potential to considerably reduce remaining risk selection incentives at the expense of a relatively low reduction of incentives for cost control.

6.1 INTRODUCTION

Many social health insurance markets are organized according to principles of regulated competition (Van de Ven et al. 2003; McGuire & Van Kleef 2018b). In such markets, the government enforces regulations to safeguard affordability and accessibility of insurance coverage, while competition among health insurers must ensure efficiency of insurance products and good quality care. To achieve affordability of basic coverage, these markets typically include premium-rate restrictions. However, a downside of such restrictions is that they create predictable profits on healthy individuals and predictable losses on the chronically ill. These predictable profits and losses confront health insurers with incentives for risk selection, which is undesirable as risk selection might violate both fairness and efficiency in health care financing (Glazer & McGuire 2000; Van Kleef et al. 2019; Van de Ven et al. 2015). Therefore, another typical feature of regulated competitive health insurance markets is risk equalization. Risk equalization compensates insurers for predictable spending variation across individuals and thereby mitigates selection incentives. However, risk-equalization formulae currently used in practice do not (yet) account for all predictable spending variation, implying that some selection incentives remain (Buchner et al. 2013; Ellis et al. 2017; Geruso et al. 2019; Van Kleef et al. 2019; Withagen-Koster et al. 2018).

The risk-equalization model applied in the Dutch health insurance system is considered to be one of the most sophisticated in the world. This model includes an extensive set of demographic, socioeconomic and morbidity-based risk adjusters. Nevertheless, it has been shown that even this model leaves considerable predictable profits and losses on selective groups. Using information from a large health survey, Van Kleef et al. (2019) found a predictable profit of around 180 euros per person per year for consumers who reported a (very) good health status in the prior year (about 75 percent of the population) and a predictable loss of around 500 euros per person per year for consumers who reported a fair or (very) poor health status in the prior year (around 25 percent of the population). In addition, the authors found predictable losses at the level of specific chronic conditions. For example, for individuals who reported to have ever suffered from diabetes, stroke, heart attack or cancer, they found predictable losses of around 130, 900, 380 and 430 euros per person per year, respectively. These predictable profits and losses might lead insurers to engage in risk selection, e.g., via the design and marketing of insurance plans.

The predictable profits and losses on the abovementioned subgroups suggest that health survey information is predictive of 'residual spending after risk equalization'

(i.e., actual spending minus predicted spending generated by the risk-equalization model). This implies that, in theory, health survey information could be used for improving the predictive power of risk-equalization models (Withagen-Koster et al. 2018; Ellis et al. 2018:77-78). However, this information is currently not used in risk-equalization models due to feasibility challenges and potential bias. For example, it will typically be considered too costly and practically infeasible to routinely collect this information for the entire population (van de Ven & Ellis 2000; Ellis et al. 2018:77-78). Health insurers, however, are not restricted by these requirements. They may use information from self-administered health surveys for their own risk assessments and in doing so, find indications of profits or losses for selective groups, creating incentives for risk selection.

A promising but understudied option to exploit the predictiveness of survey information for residual spending after risk equalization is high-risk pooling (HRP). This is a form of risk sharing between health insurers and the regulator that can be useful in mitigating remaining selection incentives (McGuire & Van Kleef 2018c). With HRP, health insurers can assign certain enrollees with high expected (residual) spending to a pool at the start of a contract period, on the basis of for instance health survey information. For enrollees in the pool, the insurer receives a compensation based on the actual (residual) spending of these enrollees once the contract period has ended (Van Barneveld et al. 1998).

While HRP can reduce selection incentives, it also comes with a price: like other forms of risk sharing, HRP reduces incentives for insurers to control costs because the compensation they receive becomes partly dependent on actual spending (Van Barneveld et al. 1996; McGuire & Van Kleef 2018c; Brammli-Greenberg et al. 2019). However, through smart choices in the design of HRP, the loss in incentives for cost control may be mitigated, e.g., by compensating insurers for high *residual* spending after the risk-equalization rather than actual spending per se. Moreover, an important advantage of HRP compared to other forms of risk sharing is that it specifically targets *predictably* unprofitable groups, leading to a more favorable tradeoff between selection and cost control.

The aim of this paper is to study the extent to which HRP based on health survey information can mitigate the risk selection incentives that remain after sophisticated risk equalization. In addition, for various HRP modalities, we assess the tradeoff between incentives for risk selection and incentives for cost control. To achieve these objectives, we simulate predictable profits and losses for selective groups under the sophisticated Dutch risk-equalization model of 2016 supplemented with

HRP. We use administrative data (N≈16.9m) on both actual and predicted spending of 2013, merged with health survey data from 2012 (N≈384k). To assess the effect of design choices on incentives for selection and cost control, we examine different pool sizes and methods for identification of pool members.

This paper is structured as follows. Section 6.2 presents an overview of design options for HRP. Section 6.3 describes the setting of the study, as well as the data and methods used. Next, Section 6.4 presents the results and Section 6.5 discusses the main findings.

6.2 HIGH-RISK POOLING DESIGN CHOICES

Risk equalization can be supplemented with risk sharing to protect insurers against large losses and to mitigate risk selection incentives. However, risk sharing also reduces incentives for cost control because payments to health insurers become (partly) dependent on actual spending. The extent to which incentives for cost control and risk selection are mitigated depends on the specific design of the risk sharing scheme. For HRP, there are three key design choices: Who is assigned to the pool and by whom? Which and how much of the spending included in HRP is compensated for? And how are these compensations financed? (Van Barneveld et al. 2001). We discuss these choices in more detail below.

6.2.1 Assignment of members to the high-risk pool

The first design choice is on the procedure of assigning individuals to the high-risk pool and the size of the pool in terms of included individuals. HRP is based on prospective assignment of individuals using information that is known at the start of the contract period (Van Barneveld et al. 2001; McGuire & Van Kleef 2018c). In theory, the assignment of members to the pool can either be done by the regulator or by health insurers themselves. However, when the regulator wants to exploit the predictiveness of (for instance) survey information for residual spending after risk equalization, assignment will most likely be done by health insurers since they typically possess (or are in a better position to obtain) this kind of information. A relevant question is then which enrollees health insurers should assign to the pool. Ideally, these are the enrollees with the highest predictable losses.

Regarding the size of the pool, the regulator can decide to use a fixed size for all health insurers or let it vary among insurers depending on differences in the risk profile of the insured population of the insurers (Van Barneveld et al. 1996, 1998). All

else equal, a larger pool size implies a greater reduction in incentives for both risk selection and cost control. It is up to the regulator to make this tradeoff.

6.2.2 Compensation for spending in the pool

The second design choice concerns the compensation for the spending of the individuals in the pool. The regulator has many options in this regard[16]. For instance, it can decide to compensate insurers for all spending of pooled members, for spending above a certain threshold, or for a certain percentage of spending (above a threshold) (Van Barneveld et al 2001). Another option is to compensate insurers on the basis of *residual* spending instead of *actual* spending. The main advantage of this is that the compensation is for those individuals with high residual spending after risk equalization (i.e., those with high losses), and not for individuals whose actual spending turns out to be well-compensated for by the risk-equalization model itself (Schillo et al. 2016).

6.2.3 Financing of the pool

The final design choice concerns the financing of the high-risk pool. This can be done externally or internally. With external financing, there is an external flow of money towards the payment system. In the case of internal financing, the high-risk pool is financed by a mandatory contribution from all health insurers. This contribution can be calculated at the end of the contract period or at the start, for example in the form of a proportional or flat reduction of the individual-level risk-equalization payments (Van Barneveld et al. 2001).

6.3 STUDY SETTING, DATA AND METHODS

6.3.1 Study setting

This study was performed in the context of the Dutch basic health insurance, which covers physician services, hospital care and prescription drugs, among other care services. The basic health insurance operates according to principles of regulated competition. This means that health insurers compete on price and quality of insurance plans, while the regulator enforces regulation to protect public objectives such as accessibility and affordability of basic coverage. These regulations include a standardized benefits package, open enrollment, community rating per insurance

16 We assume that the regulator includes all spending on all covered benefits in the high-risk pool. Of course, a regulator can also choose to only compensate for specific benefits (Van Barneveld et al. 2001). However, in contrast to including all benefits, only including specific benefits might leave some high-risk individuals unqualified for the pool. Also, not including all benefits could lead to unwanted substitution of health care or gaming (van Barneveld et al. 2001).

plan, an individual mandate to buy health insurance, and risk equalization (Van Kleef et al. 2018a).

This paper focuses on the Dutch risk-equalization model of 2016. From 2016 to 2022 this model has undergone only relatively minor changes. The Dutch risk-equalization scheme consists of three separate models: one for somatic care, one for mental care and one for out-of-pocket spending under the mandatory deductible (385 euros in 2022). This study focuses on the model for somatic care (comprising approximately 90% of total spending under the basic health insurance), which contains a broad set of socio-demographic and morbidity-based risk adjusters. The model of 2016 contains the following risk adjusters: age interacted with gender, region, socioeconomic status and source of income both interacted with age, Pharmacy-based Cost Groups (PCGs), Diagnosis-based Cost Groups (DCGs), Multiple year High Cost Groups (MHCGs), Durable Medical Equipment Cost Groups (DMECGs), yes/no morbidity[17] interacted with age, physiotherapy-spending in the prior year, geriatric rehabilitation care spending in the prior year and home care spending in the prior year (Van Kleef et al. 2018a). The risk-equalization payments under the somatic model are completely prospective and not supplemented with any risk sharing payments.

6.3.2 Data

To examine the effect of HRP and specific design choices on incentives for cost control and risk selection, we used two data sources. The first dataset contains administrative information on individual-level spending and risk adjusters for all Dutch citizens with a basic health insurance in 2013 (N≈16.9 million). These data are those that were actually used to estimate the Dutch risk-equalization model of 2016.

We merged these data with health survey data from 2012 (N≈384k) using an anonymized individual-level identification key. The health survey data contain indicators of self-reported health and lifestyle for individuals 19 years or older on September 1, 2012[18] (Public Health Monitor 2012). Specifically, these data include information on self-reported general health, nineteen chronic conditions that individuals could report to have suffered from in the last 12 months and four conditions they could report to have ever suffered from. We used these data for two purposes: 1) to predict residual spending from the perspective of a health insurer (see section 6.3.3.1 for details) and 2) to evaluate predictable profits and losses for subgroups that are potential targets of risk selection by health insurers. Our selection of groups has been extensively analyzed

17 Morbidity is defined as at least one classification in a PCG, DCG, MHCG or a DMECG.

18 Public Health Monitor (2012) of the Community Health Services, Statistics Netherlands and the National Institute for Public Health and the Environment.

in previous studies and is considered relevant when it comes to the evaluation of the Dutch risk-equalization model in terms of selection incentives (Van Kleef et al. 2013a; Van Kleef et al. 2017b; Van Kleef et al. 2019; Withagen-Koster et al. 2020).

Before conducting the analyses, we improved the representativeness of the health survey sample through a raking procedure. Via an iterative process we generated a weight for every record in our data using a set of key variables present in both the total adult population and the health survey. Application of these weights to the survey sample makes sure that the frequencies of these key variables in the health survey equal those in the population (Izrael et al. 2000; Battaglia et al. 2009). The set of key variables includes all risk-adjuster classes of the Dutch risk-equalization model of 2016, as well as 18 quantiles of mean curative somatic spending and a proxy for whether someone had died in 2013[19]. More details on the raking procedure and results on the sample's representativeness before and after raking are provided in Withagen-Koster et al. (2020).

6.3.3 Methods

Our analytical approach to evaluate the extent to which HRP can mitigate risk selection incentives and at what cost, comprised three steps. We first simulated the assignment of individuals to the pool by insurers based on health survey information. To do so, we developed a model that uses the health survey information to predict individual-level residual spending after risk equalization. Based on the predictions from this model, we assigned the top X% with the highest predicted residual spending to the pool. For the individuals in the pool, 100% of actual spending above a certain threshold is compensated. Second, we calculated the predictable profits and losses after risk equalization for selective groups identified in the health survey, with and without HRP. Lastly, using summary measures we quantified the effect of our HRP-modalities on incentives for risk selection and cost control.

The following sections describe our approach in more detail. Sections 6.3.3.1 and 6.3.3.2 start with answering the design questions raised in Section 6.2 for our specific application of HRP. Next, Section 6.3.3.3 explains how we evaluated the effect of HRP on the predictable profits/losses for selective groups. Lastly, Section 6.3.3.4 describes how we assessed the effect of HRP on incentives for risk selection and cost control.

19 Because the risk equalization model is estimated by OLS, mean spending equals mean predicted spending in the population. This is not necessarily the case for subsamples of the population, like the health survey sample. The raking procedure already corrected for some of this difference between mean spending and mean predicted spending. To set mean spending equal to mean predicted spending after raking, we multiplied the actual spending with a factor of 1.002457.

6.3.3.1 Compensation for spending included in and assignment to the high-risk pool

In this study, spending of pool members will be compensated for on the basis of residual spending above a certain threshold. This threshold was chosen such that the mean residual spending for the group of people assigned to the high-risk pool equals zero. An algorithm was used to determine the exact value of the threshold, which we determined separately for each pool size. We included 5 different pool sizes: the top-1%, top-2%, top-3%, top 4-% and the top-5% of predicted residual spending.

We assumed that assignment of individuals to the pool is done by health insurers at the start of a contract period, using health survey information. Since the compensation is based on residual spending, health insurers will want to identify and assign individuals with the highest expected residual spending to the pool. A key question is how an insurer could determine the expected residual spending. Firstly, the relationship between health survey information and residual spending must be determined, which can be done by developing a prediction model using data from a prior period. Using individual-level residual spending[20] and health survey information from a prior period, a health insurer can develop a model to predict individual-level residual spending for the upcoming contract period. This model can be developed with conventional parametric regression methods (like ordinary least squares; OLS), which are mostly used in risk equalization, but also with statistical methods that have been developed more recently, like machine learning. Prior work has shown that machine learning algorithms have the potential to better exploit the information in the data (Rose 2016) and might be better in classifying individuals with the highest predictable loss. In this study we explored both options. Specifically, we used an OLS regression with stepwise selection, as well as a random forest (RF) procedure to predict residual spending using information from the health survey. Other studies have used RF in the context of risk equalization and found that RF performed well in predicting healthcare spending (Rose 2016; Stam et al. 2021). The RF procedure creates an ensemble of many individual decision trees to protect from outliers and overfitting, with the final model being based on the aver-

20 To acquire the individual-level residual spending from the perspective of a health insurer, we followed a pragmatic approach by re-estimating the actual risk equalization model used in practice. Since a Dutch health insurer has no access to certain socio-demographic information, this replica of the risk equalization model contains the following risk adjusters: age interacted with gender, region, PCGs, DCGs, MHCGs, DMECGs, physiotherapy-spending in the prior year, geriatric rehabilitation care spending in the prior year, yes/no morbidity interacted with age and home care spending in the prior year. We estimated this model on the administrative data containing health care spending for all Dutch citizens. We then calculated the replicated individual-level residual spending by subtracting the spending predicted by the health insurer from actual spending.

age values of all estimated trees (Rose 2016). In our application, we ran 100 trees[21] with a minimum sample size of 100 in each end node.

We randomly selected 70% of the survey sample and trained both models (i.e., OLS and RF) on this selection of the data. To evaluate model performance, we tested both models on the remaining 30%. Both models yielded a prediction of the residual spending of individuals in the sample.

6.3.3.2 Financing of the pool

The high-risk pools in our analyses are financed by a reduction of the risk-equalization payment for individuals not assigned to the pool (i.e., the complementary group). In line with prior research, we choose a retrospective flat-rate contribution to ensure that the contribution is independent of the risk of enrollees (Van Barneveld 1996). We calculated this contribution as the total amount of money needed to fund the pool (i.e., the sum of actual residual spending of all individuals in the pool) divided by the number of enrollees in the complementary group. We then subtracted this fixed amount from the risk-equalization payment of every individual in the complementary group.

6.3.3.3 Predictable profits and losses for selective groups

We evaluated the effect of HRP on the incentives for risk selection and cost control using the test sample only (30% of the population, see Section 6.3.3.1). We did this to simulate how the implementation of HRP would likely occur in practice. That is, an insurer would train his prediction model on data from a prior contract period and apply it to a new contract period.

To evaluate the effect of HRP on the incentives for risk selection, we determined the predictable profits and losses for selective groups that are potential targets for risk selection, in two steps. Firstly, we determined the groups for evaluation based on the health survey information. We selected nineteen groups with specific chronic conditions individuals could report to have suffered from in the past 12 months and four groups with chronic conditions individuals could report to have ever suffered from in the past. In addition, we evaluated profits/losses for the two groups yes/no chronic condition (constructed based on the 23 groups mentioned above) and two groups based on self-reported general health: the group who reported to have a fair or (very) poor general health and the groups who reported to have a (very)

21 The optimal number of trees in a RF analysis is when the error rate does not decline any further by adding more trees. In our analyses this was the case at approximately 100 trees, which we found through trial and error.

good general health. Secondly, separately for each of five high-risk pool sizes studied in this paper, we determined the mean per person profit/loss for these groups by subtracting the actual spending from the revenue (i.e., spending predicted by the risk-equalization model plus the HRP contribution minus financing).[22]

6.3.3.4 Incentives for cost control versus. incentives for risk selection

Implementing risk sharing like HRP comes with a trade-off between incentives for cost control and incentives for risk selection. To quantify the impact of HRP on incentives for cost control, we performed a simulation in the spirit of the 'power' metric proposed by Geruso and McGuire (2016):

$$Power = 1 - \frac{1}{N}\sum_i \frac{dR_i}{dY_i}$$

where N is the number of enrollees in an insurance plan, and $\frac{dR_i}{dY_i}$ is the derivative of revenue R for person i with respect to a marginal change in spending Y for person i. This metric yields a value between 0 and 1. The higher this value, the stronger the incentives for cost control. For example, in a situation where changes in spending (dY_i) do not lead to changes in revenue (dR_i) insurers have maximum incentives for cost control, while in a situation where changes in spending result in equivalent changes in revenue, insurers have no incentives for cost control. We simulated the reduction in power due to HRP for each of the high-risk pool sizes by calculating the extra revenue a health insurer receives as a result of a 10% increase in spending[23], keeping all else equal[24].

To assess the incentives for risk selection, we calculated the weighted mean absolute result (WMAR). We first determined the mean result for each subgroup:

$$\bar{e}_g = \bar{Y}_g - R_g$$

Where $\bar{e}_g$ is the mean residual spending of subgroup g, $\bar{Y}_g$ is the mean actual spending of subgroup g and R_g is the revenue for subgroup g. The WMAR is then calculated as follows:

22 Note that the subgroups were determined on information from year t-1, while the spending, predicted spending and predictable profits/losses pertain to year t. This is also the procedure followed for risk equalization itself, where the information for the morbidity-based risk adjusters is also from year t-1.

23 Note that the percentage increase in spending level per se is irrelevant as the power metric quantifies the ratio between revenue and spending.

24 While the Dutch risk equalization model of 2016 is not supplemented with any risk sharing scheme, the model does include a risk adjuster based on spending levels in prior years which reduces the incentives for cost control (Van Kleef & Van Vliet, 2021).

$$WMAR = \frac{\sum_g (N_g * |\bar{e}_g|)}{\sum_g N_g}$$

Where N_g is the number of individuals in subgroup g and $|\bar{e}_g|$ is the absolute mean residual spending of subgroup g. A higher WMAR indicates stronger incentives for risk selection. The WMAR can be based on any selection of subgroups. Since the WMAR might be sensitive to this selection, we calculated the WMAR for four different sets of groups: 1) The group who reported not to suffer from a chronic condition and the groups of individuals who reported to have ever suffered from diabetes, stroke, heart attack or cancer, 2) all subgroups analyzed in this paper, 3) yes/no at least one chronic condition and 4) self-reported general health.

6.4 RESULTS

This section starts with a presentation of some descriptive statistics of the survey sample (Section 6.4.1), followed by results from the OLS and RF prediction models (Section 6.4.2). As we will show, the RF model performs somewhat better in identifying unprofitable individuals (i.e., those with high residual spending post risk equalization) and will therefore be used in the rest of our analysis. Based on the RF model we define five sizes of the high-risk pool. After describing these pools (Section 6.4.3), we present the per person profit/loss for selective subgroups of interest under each pool size (Section 6.4.4). Lastly, we provide insight in the tradeoff between risk selection incentives and incentives for cost control (Section 6.4.5).

6.4.1 Descriptive statistics

Table 6.1 shows descriptive statistics of the rebalanced survey sample and the in age corresponding population in the administrative data. The relative frequencies of the survey sample and the population match very well. The last two columns show the relative frequencies of the samples used to train and test the prediction models. The mean spending of the training and test sample are slightly different from that of the total survey sample and population, but the relative frequencies for the age categories and for the morbidity-based risk adjusters correspond relatively well.

6.4.2 Assigning members to the high-risk pool: OLS versus. RF

In our simulation, we assume that health insurers would assign individuals with the highest expected residual spending to the high-risk pool. A crucial question then is how insurers could determine the expected residual spending of their enrollees, assuming they have 'external' information that is not already included in the risk-equalization model (in this paper: health survey data) that they can use to predict

Table 6.1. Mean curative somatic spending and population frequencies in 2013 for selected risk-adjuster variables in the rebalanced survey sample, training sample and test sample, and the population (19 years or older on September 1, 2012).

	Survey sample (rebalanced)	Population	Training sample	Test sample
N	384,004	12,774,886	268,533	115,473
Mean spending in euros	2460	2460	2465	2448
Man, 19-34 year	11.8%	11.8%	11.7%	12.1%
Man, 35-44 year	8.7%	8.7%	8.7%	8.7%
Man, 45-54 year	9.8%	9.8%	9.7%	10.1%
Man, 55-64 year	8.4%	8.4%	8.5%	8.4%
Man, 65 year and older	10.1%	10.1%	10.1%	10.1%
Woman, 19-34 year	11.8%	11.8%	11.8%	11.6%
Woman, 35-44 year	8.8%	8.8%	8.9%	8.6%
Woman, 45-54 year	9.8%	9.8%	9.8%	9.8%
Woman, 55-64 year	8.4%	8.4%	8.5%	8.4%
Woman, 65 year and older	12.4%	12.4%	12.4%	12.3%
PCGs>0	24.1%	24.1%	24.2%	24.7%
DCGs>0	11.5%	11.5%	11.5%	11.7%
MHCGs>0	7.1%	7.1%	7.0%	7.2%
DMECGs>0	1.1%	1.1%	1.1%	1.0%
Physiotherapy spending in the previous year	2.6%	2.6%	2.6%	2.5%
Home care spending in the previous year	2.6%	2.6%	2.6%	2.6%
Geriatric rehabilitation care spending in the previous year	0.3%	0.3%	0.3%	0.2%

Note: PCGs are Pharmacy-based Cost Groups, DCGs are Diagnosis-based Cost Groups, MHCGs are Multiple year High-Cost Groups and DMECGs are Durable Medical Equipment Cost Groups.

the residual spending of their enrollees. In terms of prediction models, insurers have a universe of options. We explored two of those options: an OLS regression and a RF model. Our results show that both methods yield different results in terms of which individuals are selected for the top X% of predicted residual spending.

Figure 6.1 shows the mean actual and mean predicted residual spending for the five top percentiles of predicted residual spending for the training (panel a) and test (panel b) sample for the OLS regression. The predicted residual spending (scattered bars) and the actual residual spending (filled bars) match relatively well for the training sample. For the test sample, the predicted and actual residual spending

match less than for the training sample. This was to be expected since the model was calibrated on the training sample.

Figure 6.1. Mean per person (predicted) residual spending for the highest percentiles of predicted residual spending of the OLS prediction model for the training and test sample

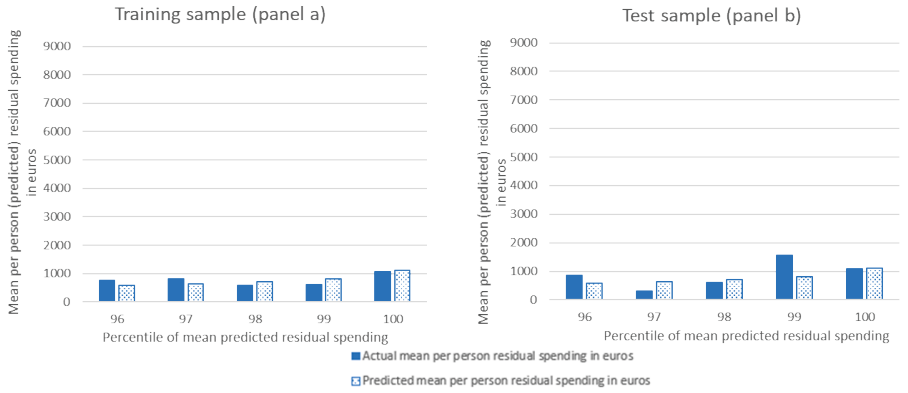
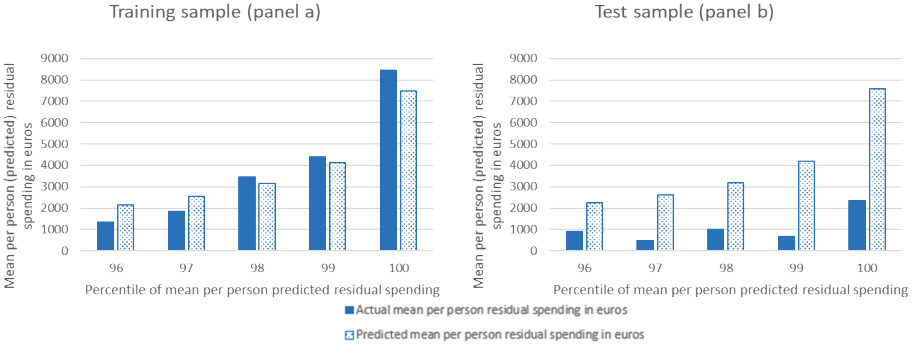


Figure 6.2 shows the same information as Figure 6.1 but then for the RF model. The predicted and actual residual spending match relatively well for the training sample (panel a) and the actual residual spending (filled bars) shows an upward trend. For the test sample (panel b), predicted and actual residual spending diverge which illustrates the problem of overfitting in the training sample. In the remainder of our analysis, we will therefore use the test sample to calculate and evaluate outcomes.

Figure 6.2. Mean per person (predicted) residual spending for the highest percentiles of predicted residual spending of the random forest model for the training and test sample.



When it comes to the assignment of individuals to the high-risk pool, the top X% of predicted residual spending identified by RF is more selective than that identified by OLS. This is illustrated in Table 6.2, which shows the mean per person actual residual spending for the top-1% to top-5% of predicted residual spending. Based on

these findings, we decided to continue our analyses with only the top X% groups as identified by the RF model.

Table 6.2. Mean per person actual residual spending in euros for the top X% of predicted residual spending under the OLS and random forest prediction model.

	Top 1%	Top 2%	Top 3%	Top 4%	Top 5%
OLS	1091*	1331*	1093*	895*	886*
Random forest	2359*	1506*	1343*	1130*	1086*

Note: Results are based on test sample (N= 115,473). An asterisk (*) means that the presented value is statistically significantly different from zero ($p<0.05$).

6.4.3 Descriptive statistics of the five high-risk pools

Table 6.3 shows descriptive statistics for each of the five high-risk pools (i.e., top-1% to top-5% of predicted residual spending based on the RF model). The table shows that with an increase of the pool size, both mean spending and mean predicted spending decrease. This could be expected as the groups become less selective and more individuals with lower (residual) spending are included (see also Table 6.2). The threshold reflects the value above which residual spending for those in the pool is reimbursed, which is chosen such that the mean residual spending of the pool becomes zero. The threshold increases with pool size because the pool becomes less selective and therefore the mean residual spending for the group included in the high-risk pool decreases (see Table 6.2). This means that – on average per person in the pool – less money is needed to reduce the mean residual spending of the pool to zero. However, more individuals are included in the pool which means that the *total* costs of financing HRP increases.

Table 6.3. Mean (predicted) spending in euros under risk equalization without a high-risk pool, the threshold for compensation and the percentage of total costs needed to finance the high-risk pool for five different pool sizes (i.e., top-1% to top-5% of predicted residual spending)

	High-risk pools based on predicted residual spending from the random forest model				
	Top 1%	Top 2%	Top 3%	Top 4%	Top 5%
Mean spending	12730	10826	10097	9571	9280
Mean spending predicted by the risk-equalization model	10371	9320	8755	8440	8194
Threshold	12961	18066	20408	23706	23807
HRP financing costs as percentage of total spending in test sample	0.96%	1.2%	1.6%	1.8%	2.2%

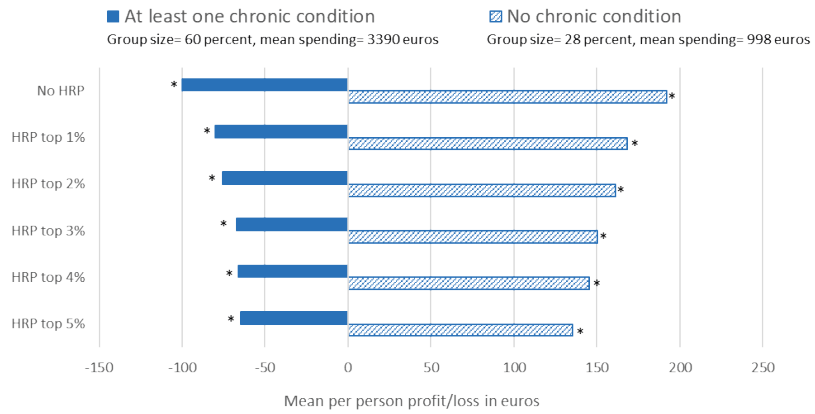
Note: all spending is in euros. Results are based on test sample (N= 115,473).

6.4.4 Mean per person profits and losses for selective subgroups under high-risk pooling

This section presents the effects of HRP on the mean per person profits/losses for selective subgroups in the survey sample. Figure 6.3 shows the mean profit/loss under the Dutch risk-equalization model with and without HRP for the individuals who reported to have suffered from at least one chronic condition (ever or in the past 12 months) and for the complementary group of individuals who reported not to have suffered from a chronic condition. As expected, the loss for the group with at least one chronic condition decreases with increasing pool size. The profit for the group without a chronic condition also decreases because of the zero-sum nature of the risk-equalization model, which means that over two complementary groups the mean profit/loss equals zero[25].

Figure 6.3 further shows that the absolute difference in compensation between the group with and the group without a chronic condition decreases from 292 euros for the situation without HRP to 200 euros for the situation with HRP for the top 5% of predicted residual spending, a reduction of 32 percent. The risk pool for the top 1% is responsible for roughly half of this reduction.

Figure 6.3. Mean per person profit/loss per year in euros under the Dutch risk-equalization model of 2016 without a high-risk pool and for 5 high-risk pools that vary in size for the groups of individuals who reported (not) to suffer from a chronic condition (ever or in the past 12 months).



Note: Results are based on test sample (N= 115,473). HRP stands for high-risk pooling. No HRP shows the mean per person profit/loss under the risk-equalization model without a high-risk pool. An asterisk (*) means that the presented value is statistically significantly different from zero ($p<0.05$). Group size does not sum to 100% over the two groups due to missing values for these groups.

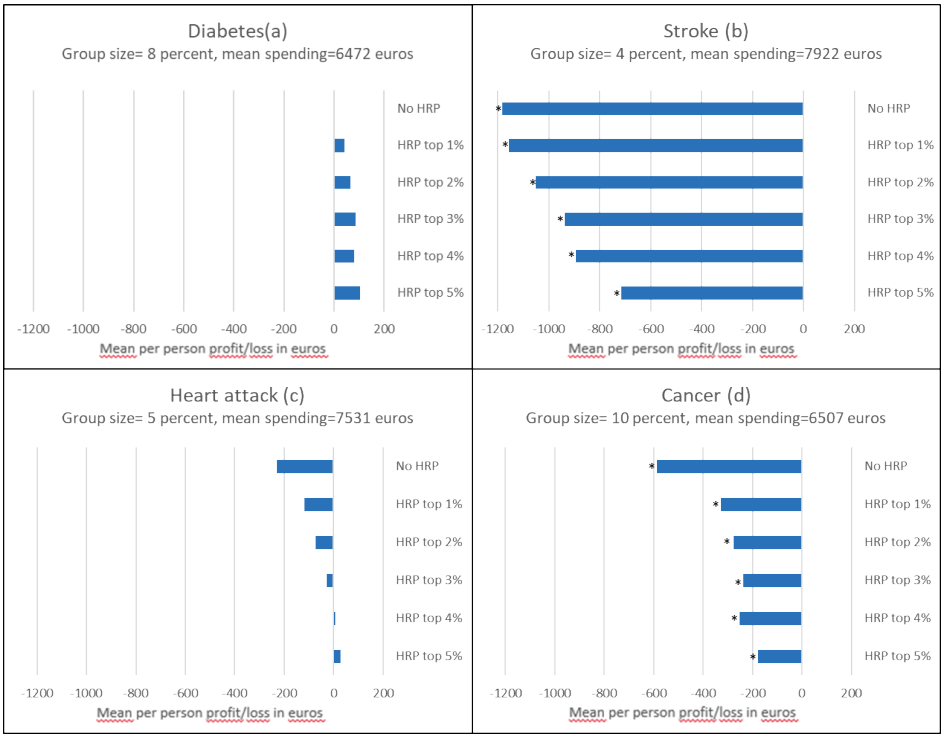
25 The mean residual spending over all these groups (including the group with missing values) is not equal to zero, despite zero sum. This is because the test sample has a mean overall residual spending of -15 euros, which is due to the inherent randomness involved in our split-sample procedure.

Figure 6.4 shows the mean per person profit/loss under the Dutch risk-equalization model with and without HRP for individuals who reported to have ever suffered from diabetes, stroke, heart attack or cancer[26]. For stroke, heart attack and cancer the loss decreases with the increase in pool size. For heart attack, the loss even turns into a profit with HRP for the top-4% and top-5% of predicted residual spending (note that for this condition, the predictable profit/loss under all high-risk pooling modalities are not significantly different from zero). For diabetes, the existing profit increases with the increase in pool size, which is related to the fact that diabetes is already well accounted for in the risk-equalization model and HRP further increases the compensation. The relative increase in profit for diabetes is largest between the scenario without HRP and HRP for the top 1% of predicted residual spending (note, however, that the profit for diabetes for these two scenarios is not statistically significantly different from zero). For the groups heart attack and cancer, the relative decrease in loss is again largest between no HRP and HRP for the top 1% of predicted residual spending (48 and 44 percent, respectively) and decreases further for both groups as the pool size increases. Lastly, for stroke, the pattern of the results differs from the other conditions: the decrease in loss reduces gradually as the pool size goes up.

Appendix A (Table A.6.1) presents mean per person profit/loss under the Dutch risk-equalization model with and without HRP for nineteen specific conditions individuals reported to have suffered from in the past 12 months. For certain conditions like heart attack, heart condition, cancer and a condition of the blood vessels, HRP strongly reduces the mean per person loss (and sometimes even turns it into a profit, like for heart condition under HRP for the top-3% to top-5%). For most groups, the largest reduction in loss can again be found between no HRP and HRP for the top 1% of predicted residual spending. The results in table A.6.1 also show that for some condition groups, like migraine, HRP hardly affects the mean per person predictable losses. The reason for this is that individuals with these conditions are most likely not assigned to the high-risk pool.

26 The mean per person profits/losses for subgroups presented in this section differ from the figures presented in the introduction of this paper based on Van Kleef et al. (2019). Even though the same health survey data are used, the results presented here are based on a selection of the health survey data (i.e., the test sample) while Van Kleef et al. (2019) used the total sample.

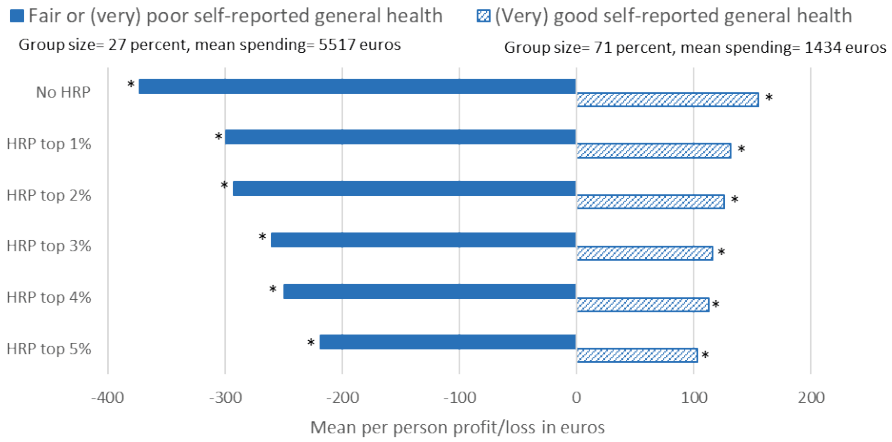
Figure 6.4. Mean per person profit/loss per year in euros under the Dutch risk-equalization model of 2016 without a high-risk pool and for 5 high-risk pools that vary in size for the groups of individuals who reported to have ever suffered from diabetes, stroke, heart attack or cancer.



Note: Results are based on test sample (N=115,473). HRP stands for high-risk pooling. No HRP shows the mean per person profit/loss under the risk-equalization model without a high-risk pool. An asterisk (*) means that the presented value is statistically significantly different from zero ($p<0.05$).

We also examined the impact of HRP on the profit/loss for groups based on a more subjective measure of health, i.e., self-reported general health. Figure 6.5 shows the profit/loss for the group who in the prior year reported a fair or (very) poor general health and for the group who reported a (very) good general health. The overall conclusions are similar to those for Figure 6.3: both the profit and loss reduce with the increase in pool size. Again, the largest change in profit/loss results from going from no HRP to HRP for the top 1% of predicted residual spending: a reduction of 15 percent for (very) good general health and a reduction of 19 percent for fair or (very) poor general health. The absolute difference between these two groups reduces from 530 euros for no HRP to 320 euros for HRP for the top 5% of predicted residual spending (i.e., -40%).

Figure 6.5. Mean per person profit/loss per year in euros under the Dutch risk-equalization model of 2016 without a high-risk pool and for 5 high-risk pools for the groups of individuals who reported a (very) good general health or a fair or (very) poor general health.



Note: Results are based on test sample (N= 115,473). HRP stands for high-risk pooling. No HRP shows the mean per person profit/loss under the risk-equalization model without a high-risk pool. An asterisk (*) means that the presented value is statistically significantly different from zero ($p<0.05$). Prevalence of groups do not add up to 100 percent due to missing values for these specific groups.

6.4.5 Incentives for risk selection versus incentives for cost control

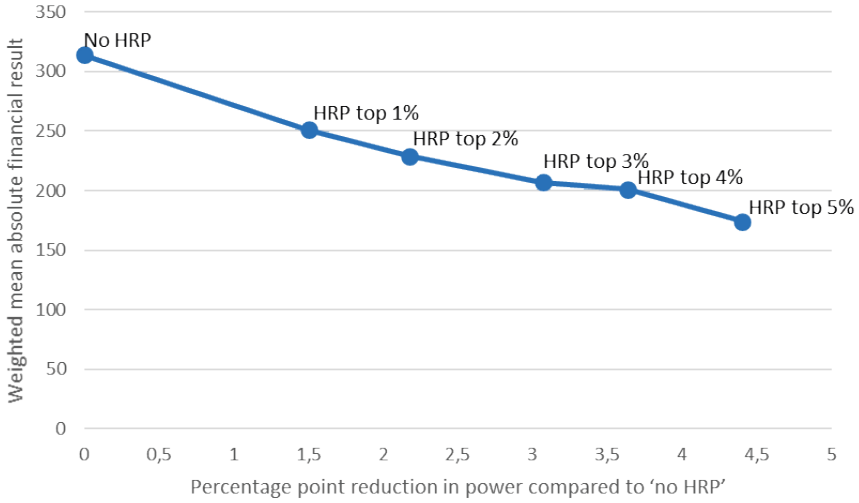
The previous section has shown that HRP can strongly reduce predictable profits/losses for specific subgroups, thus reducing the incentives for insurers to engage in risk selection. As described, however, like any other form of risk sharing HRP also reduces the incentives for cost control. This section examines the tradeoff between incentives for cost control and risk selection under our five high-risk pool sizes. To indicate the effects of HRP on the incentives for cost control, we use the Power-measure (see Section 6.3.3.4), which ranges from 0 (i.e., no incentives for cost control as payments are based fully on actual spending) to 1 (i.e., full incentives for cost control as payments are completely independent of actual spending). Specifically, we simulated the change in Power for each of the high-risk pool sizes relative to no HRP.

Figure 6.6 plots the percentage point reduction in Power compared to no HRP for the different HRP modalities against the weighted mean absolute result (WMAR); a higher WMAR indicates stronger risk selection incentives. For the WMAR in Figure 6.6 we included the subgroups no self-reported chronic condition and the 4 subgroups individuals reported to have ever suffered from: diabetes, stroke, heart attack or cancer (see Section 6.3.3.4).

Figure 6.6 shows that, as expected, the incentives for both cost control and risk selection decrease with increasing pool size. The largest relative decrease in selection incentives is achieved under HRP for the top 1% of predicted residual spending (decrease of 16 percent). This comes at the cost of a 1.5 percentage point decrease in incentives for cost control (relative to no HRP). The risk selection incentives continue to decrease as the pool size increases, eventually resulting in a total decrease in incentives for risk selection (relative to no HRP) of 43 percent under HRP for the top 5% of predicted residual spending. Under this modality, the total decrease in incentives for costs control is 4.4 percentage points.

To examine the sensitivity of WMAR to the selection of subgroups used, we also

Figure 6.6. Risk selection incentives (WMAR) plotted against the percentage point reduction in incentives for cost control (Power).



Note: HRP stands for high-risk pooling. No HRP refers to the situation of risk equalization without a high-risk pool. For the calculation of the weighted mean absolute result (WMAR) the groups no self-reported chronic condition and the groups individuals reported to have ever suffered from diabetes, stroke, heart attack or cancer have been included. See Section 6.3.3.4 for the formulas of the Power-measure and WMAR.

calculated the WMAR using other selections of subgroups, namely 1) all subgroups analyzed in this paper, 2) yes/no at least one chronic condition and 3) self-reported general health. Depending on the selection, the reduction in WMAR (i.e., in risk selection incentives) ranges between 33 and 45 percent under HRP for the top 5% of predicted residual spending relative to no HRP (results not shown).

6.5 DISCUSSION

In this paper we investigated the potential effects of High-Risk Pooling (HRP) as a supplement to sophisticated risk equalization. A crucial feature of HRP is that it allows for ex-ante assignment of high-risk individuals to the pool, which in our case is done by insurers based on health survey information. In a simulation on data from the Netherlands we examined the effects of HRP on risk selection incentives and incentives for cost control. In a first step, we identified candidates for the pool using a RF model predicting residual spending post risk equalization. We compared identification of high-risk individuals by RF to identification by OLS and found that RF was better able to identify these high-risk individuals than OLS. In a next step, we applied five different pool sizes (i.e., the top-1%, top-2%, top-3%, top-4% and top-5% of the distribution of predicted residual spending respectively) and calculated the mean per person profit/loss for specific subgroups identifiable in the health survey data and compared the results to the situation without HRP. In addition, for each of the pool sizes we evaluated the reduction in incentives for cost control. Our findings suggest that our HRP modality can lead to a considerable reduction in remaining selection incentives by sacrificing a relatively low share of incentives for cost control.

In our simulations, selection incentives (i.e., predictable profits and losses) gradually reduce with a larger pool size. We found the largest marginal reduction in selection incentives when moving from no high-risk pool to HRP for the top 1% individuals with the highest predicted residual spending. When moving from HRP for the top 1% to larger pool sizes (up to the top 5%) the marginal reduction in selection incentives goes down. The reason for this is that our prediction model identifies a more selective group in terms of unprofitable individuals in the 100th percentile of predicted residual spending compared to the 96th to 99th percentiles.

Another finding is that for some chronic conditions HRP reduces profits/losses substantially (e.g., those who suffered from cancer) while for other groups the effect is moderate (e.g., those who ever suffered a stroke) or absent (e.g., those who suffered from migraine in the past 12 months). The reason for this is to be found in the underlying prediction model. The RF procedure only selected the conditions relevant for identifying individuals with high residual spending to include as indicators in the model. Therefore, conditions that are included in the prediction model will be better represented in the top 5% of predicted residual spending and thus benefit more from HRP than conditions that are not included in the prediction model.

Since HRP (as a form of risk sharing) inherently reduces incentives for insurers to control costs (Van Barneveld et al. 1996; McGuire & Van Kleef 2018c; Brammli-Greenberg et al. 2019), regulators considering the use of HRP are faced with a tradeoff between risk selection and cost control. Depending on how regulators weigh the relative importance of 'risk selection' and 'cost control', however, the reduction in selection incentives could outweigh the reduction in incentives for cost control. Although our study ignored such normative weighing, our results do suggest that a relatively small decrease in incentives for cost control can go a long way in reducing remaining selection incentives.

A key assumption underlying this study is that health insurers have additional individual-level information to predict spending that is not included in the risk-equalization model itself. In addition to health survey information, which depending on the context might be challenging to acquire (Ellis et al. 2018), other potential data sources that might be used for HRP are administrative data, multiyear diagnostic data, and diagnostic information from general practitioners (Van Veen et al. 2015a; Van Kleef et al. 2018b). Like the survey information used in this study, these other types of data are also expected to have predictive power regarding residual spending. We do not expect, however, that these other data sources are perfect substitutes for health survey information since the latter includes subjective expectations about future healthcare spending that cannot be picked up (fully) by administrative and diagnostic data. Instead, we expect that health survey information and these other data sources might be complementary when it comes to the assignment of high-risk individuals to the pool. The effects of using these different data sources together for the purpose of HRP remains an empirical question. In addition to alternative data sources, insurers might also use other machine learning techniques, like penalized regression or a super learner (Rose 2016), depending on the available data. Combined, using more data and/or other machine learning techniques to identify high-risk individuals might lead to the identification of a more selective group, which when used for HRP might result in a more favorable tradeoff between selection and cost control (given an equal reduction in incentives for cost control). More research is needed to evaluate these options.

To our knowledge, this is the first study to investigate health survey information to prospectively assign enrollees to a high-risk pool in the context of sophisticated risk equalization. Another strength of this study is that we had access to rich administrative data for all Dutch citizens (N≈16.9m) that we could combine with unique data from a large health survey (N≈384k). Also, this study contributes to the discussion of using machine learning techniques in risk equalization. Nevertheless, our study

comes with at least two limitations. Firstly, our findings based on the health survey data are conditional on the adult Dutch population of 19 years and older. It is possible that the incentive effects of HRP would be different when considering the total population. Secondly, in our simulations of incentives for cost control using the Power metric, we did not include the effect of risk adjusters included in the risk-equalization model. Since the Dutch risk-equalization model includes risk adjusters based on prior year spending (e.g., MYHCG), incentives for cost control are not 100% (i.e. Power<1) without HRP (Van Kleef & Van Vliet 2022). Also, the Power metric only measures the direct (negative) effect of HRP for incentives for cost control. In addition to this direct (negative) effect, there might be indirect (positive) effects on incentives for cost control. For example, a decrease of selection incentives reduces the expected returns on investments in risk selection, which might lead insurers to (further) shift their investments/attention to other strategies for cost control such as improving the efficiency of care. More generally, a decrease of risk selection incentives is expected to increase efficiency as an outcome of competitive health insurance markets (Van de Ven et al., 2021).

In this study we examined the use of health survey information to assign individuals to a high-risk pool. However, the practical implementation of HRP in general needs more work. Firstly, HRP comes with several design options which can be chosen such to best fit the specific context in which it is applied. The extent to which alternative designs affect incentives and which design choices yield the optimal tradeoff between selection and cost control is an empirical question and requires further research. As this combination of design choices crucially depends on the weight attached to risk selection and cost control by a regulator, it is important that regulators specify these weights. Secondly, to protect the level-playing field between health insurers, a regulator might need to regulate the administration of the health survey or even supply health insurers with the (results of the) same prediction model to assign individuals to the pool. The reason is that not all health insurers may have the capacity to develop an equally performing prediction model, for instance because of a small portfolio.

Relative to many other possible forms of risk sharing, including proportional risk sharing, reinsurance, and risk corridors (McGuire & Van Kleef 2018c), an important advantage of HRP is that risk sharing funds are explicitly targeted at *predictably* unprofitable groups, which is an important property in the light of the selection-cost control tradeoff; other forms of risk sharing also compensate for *unpredictably* unprofitable groups, leading to a less favorable selection-cost control tradeoff (Van Barneveld 2001; McGuire & Van Kleef 2018c). In addition, HRP enables health insur-

ers to use information predictive of (residual) health care spending that is not used in risk equalization to improve compensation, which they otherwise might use for risk selection purposes. In that light, we believe HRP is a promising option to mitigate risk selection incentives under sophisticated but imperfect risk equalization.

APPENDIX A. Mean spending, mean predicted spending and mean per person profit/loss for selective groups identifiable in the health survey under the Dutch risk-equalization model and under five different HRP modalities

Table A.6.1 Mean spending, mean predicted spending and mean financial result for selective subgroups identifiable in the health survey under the Dutch risk-equalization model and under five different HRP modalities

Groups	Group size	Mean spending	Financial result under different HRP modalities					
			No high-risk pool	Top 1%	Top 2%	Top 3%	Top 4%	Top 5%
Stroke	1%	9895	-1276*	-1212*	-1144*	-1022*	-971*	-973*
Heart attack	1%	9312	-1073	-720	-706	-621	-373	-358
Heart condition	3%	8369	-374	-270	-203	50	98	150
Cancer	3%	10419	-1161*	-158	-161	-19	-125	-68
Migraine	12%	2380	-113*	-121*	-124*	-114*	-114*	-120*
Blood pressure	22%	4372	-103*	-96	-72	-32	-32	-23
Blood vessels	4%	8264	-1233*	-1131*	-942*	-714*	-717*	-660*
Asthma	9%	4725	-198*	-176*	-157	-110	-87	-79
Psoriasis	3%	3627	-282	-293*	-256	-256	-235	-225
Eczema	4%	2832	-73	-91	-90	-92	-98	-92
Severe/recurring dizziness	4%	6128	-755*	-676*	-682*	-655*	-616*	-615*
Severe/recurring disease of intestines	4%	5334	-672*	-658*	-645*	-588*	-569*	-565*
Incontinence	8%	5814	-236*	-198*	-188*	-150	-148	-146
Wear of joint	19%	4855	-295*	-277*	-265*	-261*	-259*	-253*
Joint inflammation	6%	5818	-320*	-298*	-301*	-285*	-265*	-263*
Severe/recurring condition of back	11%	4135	-208*	-182*	-184*	-178*	-183*	-178*
Severe/recurring condition of neck	10%	3871	-219*	-184*	-189*	-189*	-187*	-186*
Severe/recurring condition of elbow	7%	4531	-98	-83	-62	-44	-32	-33
Other chronic condition	15%	4948	-347*	-262*	-259*	-220*	-202*	-192*

Note: An asterisk (*) indicates if the mean residual spending is statistically significant different from zero ($p<0.05$)

7

Conclusion

7. CONCLUSION

This dissertation has demonstrated how health survey information can be used in the evaluation and improvement of risk equalization. To that end, chapters two to six have answered five research questions, of which the first three focused on the use of health survey information for the evaluation of risk equalization and the last two focused on the use of health survey information to improve risk equalization (i.e., reduce predictable profits and losses and thereby selection incentives). The results of each of these studies show that health survey information is useful for both the evaluation and improvement of risk equalization. Below, Section 7.1 summarizes the findings of chapter two, three and four and Section 7.2 summarizes the findings of chapter five and six.

7.1 Evaluating risk equalization using health survey information

Chapter two focused on the extent to which health survey information can explain variation in residual spending (i.e., actual spending minus predicted spending by the risk equalization model) in the Dutch health insurance market. We developed a prediction model with predictors based on information from the health survey to explain and predict individual-level residual spending. We also examined the correlation between *actual* and *predicted* individual-level residual spending within groups and calculated the absolute value of predicted residual spending per group. The results show that the health survey information was able to explain a small but non-negligible part of the variation in residual spending in both the morbidity and non-morbidity group as indicated by the risk-equalization model (i.e., yes/no at least one morbidity-based risk adjuster flag). Moreover, the prediction model was able to explain more of the variation in residual spending in groups with high predicted spending by the risk-equalization model, suggesting larger predictable losses for the individuals identified by the risk-equalization model as being high risks. In addition, we also found that those with the highest predicted residual spending (i.e., highest profits/losses) are also identified by the risk-equalization model as the highest-risk individuals (i.e., those with the highest predicted spending). This shows that while the risk-equalization model correctly identifies high-risk individuals, the spending predicted by that model is not high enough.

In chapter three we used the health survey information to evaluate predictable profits and losses over multiple consecutive contract periods. The relevance of this exercise is that individuals do not necessarily switch health plans at the end of every contract period, but instead might stay enrolled with the same health plan for a longer time. Under this assumption, chapter three studied how the predictable

profits and losses for selective groups identified in the health survey information develop over subsequent contract periods. More specifically, we identified selective groups of healthy and chronically ill in the health survey in year t and examined actual spending, predicted spending generated by the risk-equalization model, and the predictable profits and losses for these groups for year t+1, t+2 and t+3. While we expected to find an increase in both actual spending and predicted spending within subgroups, the results show that actual and predicted spending remain quite stable for all researched groups. As a result, healthy subgroups remain predictably profitable over time and unhealthy subgroups remain predictably unprofitable over time. These results show that to the extent that enrollees stay enrolled with the same health plan for multiple contract periods, risk-selection incentives are stronger than suggested by empirical evaluations based on one contract period.

Chapter four examined a possible driver of remaining predictable profits and losses for selective groups identified in the health survey: end-of-life spending. To study to what extent these profits and losses are driven by end-of-life spending, we simulated the predictable profits and losses with and without end-of-life spending while correcting for the overall difference in mean spending between these two situations. The results show that for some chronic conditions (like stroke and cancer) end-of-life spending significantly contributes to remaining predictable losses. For other groups we did not find a significant contribution of end-of-life spending. These findings suggest that for specific groups predictable losses might be further reduced by better correcting for risk heterogeneity related to end-of-life spending.

7.2 Improving risk equalization using health survey information

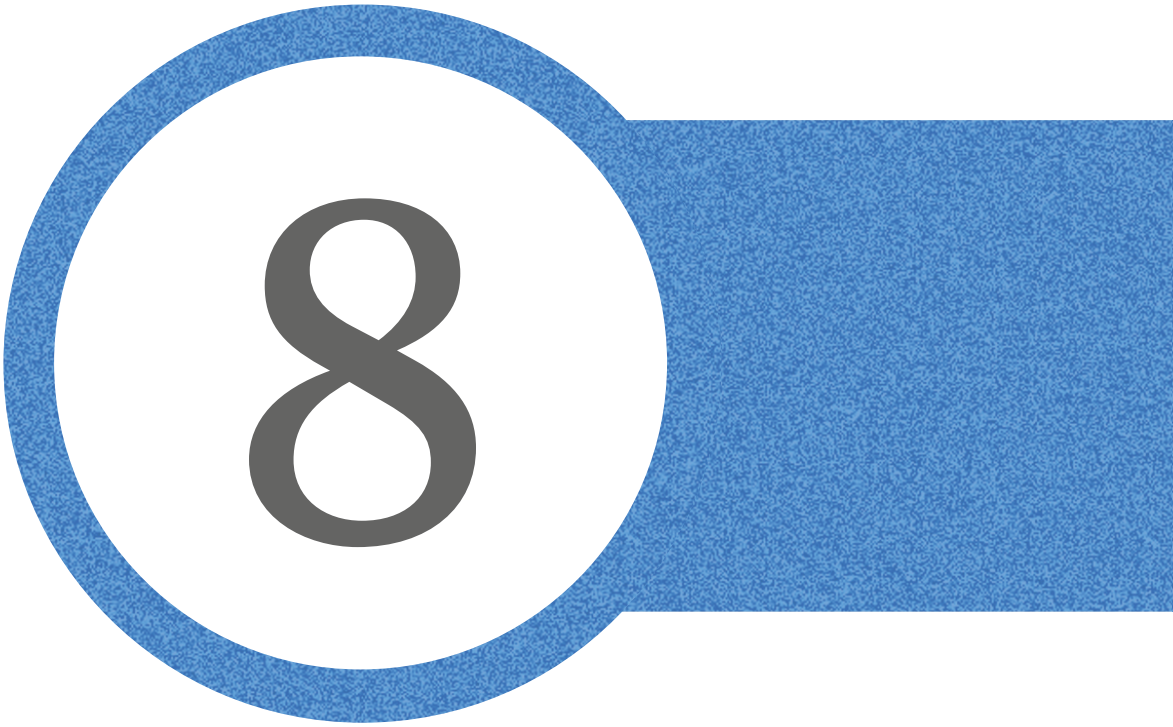
Chapters two, three and four indicate that health survey information is predictive of residual spending post risk equalization. However, including this information *directly* into the risk equalization model is often not possible. Chapter five and six have therefore researched alternative strategies to use health survey information in the risk-equalization system and evaluated the effect thereof on the predictable profits and losses for selective groups identified in the health survey information.

Chapter five focused on the option to exploit health survey information in risk equalization through constrained regression. This is a method in which the payment weights generated by the risk-equalization model are modified conditional on predefined constraints, which in this study were based on self-reported general health. An important precondition when basing constraints on information only available for a subgroup of the population (in this study health survey data), is that these data are representative for the entire population. While the health survey data

were already quite representative for the population, we further improved upon this by applying a raking procedure before using these data as a basis for the constraints. Next, we developed and evaluated five constrained-regression models in which the existing profits/losses for the groups fair or (very) poor self-reported general health and (very) good self-reported general health were reduced by 20, 40, 60, 80 and 100 percent. We compared the profits/losses for selective groups of chronically ill and healthy individuals under each of the five constrained-regression models to the profits/losses for these groups under a risk-equalization model without constraints. The results show that a relatively light constraint (i.e., 20 to 40 percent reduction of the predictable profit/loss on groups defined by self-reported health) can already improve risk equalization in terms of reduced profits/losses for various groups. The results also show, however, that a heavy constraint (i.e., 80 to 100 percent reduction of the predictable profit/loss on groups defined by self-reported general health) can perform worse in terms of profits/losses for selective groups compared to risk equalization without constraints. In line with prior research, we also found that constrained regression comes with a price: while it reduces profits/losses for some groups, it increases profits/losses for other groups (Van Kleef et al. 2017b). However, using a standardized metric which summarizes the predictable profits and losses across sets of relevant groups, we conclude that the improvement for some groups can outweigh the deterioration for others.

Chapter six researched the option to exploit the predictiveness of health survey information by supplementing the risk-equalization model with an innovative form of high-risk pooling. In contrast to a regulator, a health insurer is not restricted in the use of health survey information for its own risk assessments. By allowing health insurers to assign members to a high-risk pool at the start of the contract period and compensating health insurers for these pooled members at the end of the contract period, selection incentives may be mitigated. To examine to what extent high-risk pooling based on health survey information can mitigate such incentives, we first simulated the assignment of members to the pool from the perspective of a health insurer. To that end, we used the health survey data to develop a prediction model to identify individuals with high expected residual spending after risk equalization and assigned those with the highest predicted residual spending to the pool. Compared to the regulator, a health insurer might use different and/or more advanced identification techniques. Therefore, we compared identification of high-risk individuals by an ordinary least squares (OLS) model (i.e., as used by the regulator) to identification by a random forest model (i.e., a more advanced technique a health insurer might use) and found that the random forest model was better able to identify high-risk individuals in our data. Next, we calculated the predictable profits

and losses for selective groups identified in the health survey for a risk-equalization model supplemented with five different high-risk pool sizes (i.e., the top-1%, top-2%, top-3%, top-4% and top-5% of the distribution of predicted residual spending, respectively) and compared these to the predictable profits and losses under the same risk-equalization model but then without high-risk pooling. We found that implementing a high-risk pool for just the top 1% of predicted residual spending can already mitigate a considerable share of remaining predictable profits/losses. Lastly, we quantified the tradeoff between incentives for risk selection and incentives for cost control. Our results show that our innovative form of high-risk pooling has the potential to mitigate a relatively large share of remaining selection incentives at the expense of a relatively low reduction of incentives for cost control.


8

Discussion

8. DISCUSSION

This chapter discusses the findings of this dissertation. Section 8.1 first discusses the use of health survey information in risk equalization, and Section 8.2 highlights some additional points regarding variation in (residual) spending and risk selection incentives. Section 8.3.1 discusses several options to further improve risk equalization, followed by Section 8.3.2 that specifically discusses the use of constrained regression and high-risk pooling as innovative methods to reduce risk selection incentives. Section 8.4 recommends regulators to make explicit normative choices regarding tradeoffs related to improvement of (sophisticated) risk equalization, followed by Section 8.5 which closes with the international relevance.

8.1 Health survey information and risk equalization

This dissertation has investigated several promising options for exploiting health survey information in the evaluation and improvement of risk equalization. Using such a large and rich health survey comes with at least two important advantages in this context. First, the outcomes of studies using information from this survey are surrounded with less statistical uncertainty compared to studies that had to resort to smaller health surveys (Van Kleef et al. 2017a). Second and related to the first advantage, the size of this health survey also provides more statistical power to evaluate and research strategies to improve risk equalization compared to smaller health surveys.

Evaluating and improving risk equalization requires data not explicitly included in risk equalization. Obviously, apart from health survey information, there are other external data sources that can be used for this purpose, such as claims data, multiyear diagnostic data, and diagnostic information from general practitioners (Van Veen et al. 2015a; Van Kleef et al. 2018b). For some of the applications of health survey information as researched in this dissertation, these other data sources might be more fitting. For instance, high-risk pooling might be more effective in identifying high-risk individuals when using claims data. Even so, we believe health survey information can still be useful. As Ellis et al. (2018) have stated, health survey information is mostly useful to improve compensation for specific groups of concern instead of improving the overall fit of the risk-equalization model. Another important advantage of health survey data is that it includes information regarding the subjective health of respondents (e.g., how individuals perceive their own health status and health care use), which other data sources are unlikely to fully capture. The extent to which health survey information holds additional value in evaluating

and improving risk equalization compared to these other data sources is an empirical question, requiring additional research.

Using health survey information for the evaluation and improvement of risk equalization comes with several caveats. Firstly, since health survey information is not available for the entire population, it is important that this subpopulation is representative for the population, like the health survey used in this dissertation. While this is important for the generalizability of study outcomes, it is also crucial for the use of health survey data in strategies to improve risk equalization, like constrained regression. Basing the constraints on a poorly representative subpopulation could bias the outcomes (i.e., payment weights). Secondly, while the inherently subjective nature of health survey information has its advantages as mentioned above, it remains subjective and therefore dependent on the truthful answers of respondents.

8.2 Variation in residual spending and risk selection incentives

Chapters two, three and four have shown that while the sophisticated Dutch risk-equalization model already compensates insurers for a large share of predictable spending variation among individuals, health survey data still contain information with additional explanatory power. This is in line with conclusions from previous research (Ellis et al. 2013a; Stam et al. 2010). We concluded that the Dutch risk equalization model (still) does not always correctly identify the risk profile of individuals. More specifically, chronically ill people do not always have a morbidity flag and for those who do, the spending predicted by the risk-equalization model is not always high enough. Moreover, we have shown that this problem persists over time. For some of the subgroups identified in the health survey we found end-of-life spending to be an important driver of predictable profits and losses, but not for others. More research into other possible drivers of predictable profits and losses (e.g., lifestyle or the interaction with mental health care) is necessary to further mitigate risk selection incentives, especially considering that these incentives might have been underestimated so far.

While chapter two showed that health survey information can explain variation in residual spending, it must be noted that the prediction model we developed did not only contain self-reported chronic conditions, but also lifestyle indicators. Examples of such indicators are smoking behavior and alcohol consumption. Unlike age, gender and health status, lifestyle indicators are examples of risk factors for which a regulator typically does not desire solidarity and for that reason are not included in the risk-equalization model as risk adjusters (Van de Ven & Ellis 2000). However, given that lifestyle indicators are known to have predictive power for medical

spending, it would still be interesting for further research to examine the extent to which self-reported lifestyle indicators could explain variation in residual spending. The reason is that health insurers might use lifestyle indicators to risk select their enrollees, for example by offering attractive supplementary health insurance to members of a sports club. In addition to lifestyle indicators, the prediction model developed and applied in chapter two also contained self-reported mental health indicators. While the Dutch risk equalization scheme includes a separate model for mental health care use, no such indicators are included in the somatic model. Since the mental health state of individuals might affect their overall health (i.e., the interaction between somatic and mental health) and therefore affect healthcare use and spending, researching the extent to which these indicators explain variation in residual spending might also be interesting.

Chapter three specifically examined the persistence of predictable profits and losses for selective groups identifiable in the health survey information over time. The results showed that the predictable profits and losses for these groups remained quite stable over three consecutive years. While the per-year predictable profits/losses are similar to studies evaluating risk-selection incentives based on one contract period (Van Kleef et al. 2017b; Van Kleef et al. 2019), the relevance of the findings of chapter three is that risk-selection incentives based on one contract period might underestimate the true incentives for risk selection, especially for chronically ill individuals. The reason is that because chronically ill are already less inclined to switch health plans compared to healthy individuals (Atherly et al. 2020; Boonen et al. 2016; Van der Schors et al. 2020), chronically ill are more likely to stay enrolled in a certain health plan for more than one contract period, confronting health insurers with their unprofitability for multiple contract periods. This might reinforce the reluctance of health insurers to actively invest in the quality of health care for unprofitable groups of chronically ill (KPMG 2014; 2020)

8.3.1 Guiding improvement of the risk equalization model

To mitigate risk-selection incentives, it is important to improve the risk equalization model. As explained in chapter one, there are different strategies to do so. One option is to improve the risk-equalization model by adding new risk adjusters or to refine existing risk adjusters. This will result in more homogenous risk groups and enable more accurate predictions of spending. To guide such improvement, chapter four examined to what extent end-of-life spending drives remaining predictable profits and losses. The results showed that for some chronic condition groups, end-of-life spending significantly contributes to predictable losses. To reduce selection incentives, better accounting for end-of-life spending in risk equalization might

reduce the predictable profits and losses for these groups. However, since a large share of ex-post spending on the deceased can be explained by ex-ante spending on the sick (Einav et al. 2018), better incorporating the severity of chronic conditions might already compensate for (some of the) end-of-life spending related variation. In addition, doing this might also reduce unpriced risk heterogeneity within morbidity groups included in the risk-equalization model (as identified in chapter two). A potential strategy to better incorporate the severity of chronic conditions can be found in the risk-equalization system used in the Medicare program in the United States, in which Hierarchical Condition Categories (HCC) are used. This system explicitly accounts for the severity of chronic conditions. However, implementing such a system requires the availability of ICD-information, which is currently not the case in the Netherlands. To what extent better incorporating the severity of chronic conditions might compensate for end-of-life spending remains an empirical question, requiring further research.

A relevant question, however, is to what extent risk equalization can still be improved by adding new or refined risk adjusters, especially considering that new risk adjusters that have been added to the Dutch model in recent years seem to have had a decreasing marginal contribution in explaining variation in medical spending post risk equalization (Eijkenaar & Van Vliet 2017; Eijkenaar et al. 2017; Van Kleef & Van Vliet 2022). Chapter two has shown that health survey information is able to explain only a relatively small share of the predictable spending variation post risk equalization. This suggests that the potential for improving the risk-equalization model by adding more or more refined risk adjusters might be limited, at least with the data sources currently available. To the extent that additional data sources are not available or useful, other strategies to improve risk equalization are required to further mitigate risk selection incentives. As illustrated in chapter five and six, constrained regression and high-risk pooling are promising examples of such strategies.

8.3.2 Alternative innovative methods to reduce selection incentives: constrained regression and high-risk pooling

Chapter five and six have researched the potential of constrained regression and high-risk pooling, respectively, to improve risk equalization and showed that these methods can successfully reduce remaining predictable profits and losses. Although not studied in this dissertation, these options might prove valuable complements. The reason is that the two methods redistribute money differently and can target different subpopulations. Constrained regression transfers money through the risk-adjusters included in the risk-equalization model and influences the payment weights of the model itself. Depending on the constraint, constrained regression

can target a very broad or very specific group (e.g., a specific chronic condition) anywhere in the (residual) spending distribution. High-risk pooling on the other hand, transfers money from low-risk individuals to high-risk individuals at the end of a contract period (i.e., the group not included in the high-risk pool to the group included in the high-risk pool), targeting the group of individuals concentrated at the end of the (residual) spending distribution. High-risk pooling could then for example reduce the remaining predictable profits and losses left in the market after constrained regression. The extent to which such a combined approach can be successful and what the implications are for the tradeoffs accompanying both methods (see below), are empirical questions which further research should examine. The methodological approaches developed and applied in chapter five and six to quantify the tradeoffs could prove useful in this regard.

More research in general into constrained regression and high-risk pooling is also needed. For high-risk pooling, further research could examine different design choices (including different statistical techniques for identification of high-risk individuals) and their effect on the selection/cost-control tradeoff. For constrained regression, more research into group selection as a basis for the constraints is necessary. For instance, to what extent is there a difference in predictable profits/losses when basing the constraints on general groups (like in chapter five) compared to basing them on very specific and heavily underpaid groups? Lastly, for both constrained regression and high-risk pooling it will be interesting to study the results in terms of predictable profits and losses using data other than health survey information.

As mentioned, the use of constrained regression and high-risk pooling comes with a tradeoff. Specifically, constrained regression involves a tradeoff between improved compensation for some groups and deteriorated compensation for others, and high-risk pooling involves a trade-off between incentives for risk selection and incentives for cost control. In chapters five and six we attempted to quantify the relevant tradeoff. Given the choices we made in these chapters, we found that for constrained regression the improvement for some groups can outweigh the deterioration for other groups, and for high-risk pooling that the reduction in selection incentives can outweigh the reduction in incentives for cost control (given normative social preferences and priorities). We therefore concluded that both constrained regression and high-risk pooling based on health survey information are promising options for improving risk equalization. However, accurately assessing the tradeoffs posed by constrained regression and high-risk pooling requires normative choices regarding (i) how improved compensation for some groups are weighted relative to

deteriorated compensation for other groups, and (ii) how incentives for risk selection are gauged versus incentives for cost control.

8.4 Normative choices required

Assessing alternative strategies to improve risk equalization requires regulators to make explicit normative choices. The reason is that these strategies, including constrained regression and high-risk pooling, typically present the regulator with a tradeoff. To properly evaluate and refine such strategies requires the regulator to explicitly weigh the welfare effects of for instance risk selection versus cost control in the health insurance system. Furthermore, the regulator should specify for which groups reducing predictable profits/losses is most important. When it comes to the evaluation of risk equalization, reducing predictable profits/loss might be more important for some groups than for others. Since health insurers in the Netherlands are currently not incentivized to invest in good quality health care for chronically ill (KPMG 2014; 2020), the regulator could, for instance, explicitly state that reducing predictable losses for (specific groups of) chronically ill is more important than reducing predictable profits for healthy individuals. Another aspect the regulator should specify is whether and to what extent some forms of risk selection are considered worse than other forms of risk selection, for instance deliberately not contracting good quality health care versus attracting healthy individuals via sales and marketing. Once a regulator has made such explicit normative choices, researchers can more effectively work on approaches to quantify the tradeoffs related to alternative strategies to improve risk equalization. The approaches developed in chapters five and six have deliberately left room to incorporate differentiated weighting depending on how a regulator values the elements involved in the tradeoff.

8.5 International relevance

All studies in this dissertation have used data from the Netherlands, making the conclusions directly relevant for the Dutch context. For two reasons, we believe that our findings are also relevant for other countries with regulated competitive health insurance markets. Firstly, like the Dutch model other risk-equalization models used around the world mostly use administrative data as a basis for risk adjusters (McGuire & Van Kleef 2018a). Therefore, data not explicitly included in risk equalization, like health survey data, might be useful in these other markets as well. Secondly, these other models too are known to undercompensate insurers for chronically ill people (e.g., Bauhoff 2012; Buchner et al. 2013; Carey 2017a; Ellis et al. 2013b; Ellis et al. 2017; Geruzo et al. 2019; McGuire et al. 2013; Shmueli & Nissan-Engelcin 2013). This dissertation might provide directions for improving the evaluation and design of risk-equalization systems also in these other countries.


R

References

Abaluck, J., Gruber, J. (2011) Choice inconsistencies among the elderly: evidence from plan choice in the Medicare part D program. American Economic Review 101(4): 1180-1210. Doi: 10.1257/aer.101.4.1180

Atherly, A., Feldman, R.D., Dowd, B., Broek-Altenburg, E. van den (2020). Switching costs in medicare advantage. Forum for Health Economics and Policy 23 (1). Doi: 10.1515/fhep-2019-0023

Barneveld, E.M., Lamers, L.M., van Vliet, R.C.J.A., van de Ven, W.P.M.M. (1998). Mandatory pooling as a supplement to riskadjusted capitation payment in a competitive health insurance market. Social Science Medicine 47(2):223-232.

Barneveld, E.M., Lamers, L.M., van Vliet, R.C.J.A., van de Ven, W.P.M.M. (2001). Risk sharing as supplement to imperfect capitation: a tradeoff between selection and efficiency. Journal of Health Economics 20: 147-168.

Barneveld, E.M., van Vliet, R.C.J.A., van de Ven, W.P.M.M. (1996). Mandatory high-risk pooling: an approach to reducing incentives for cream skimming. Inquiry 33(2):133-143.

Battaglia, M.P., Hoaglin, D.C., Frankel, M.R. (2009). Practical considerations in raking survey data. Survey practice 2 (5).

Bauhoff, S. (2012). Do health plans risk-select? An audit study on Germany's Social Health Insurance. Journal of Public Economics 96 (9-10), 750–759.

Boonen, L.H.H.M., Laske-Aldershof, T., Schut, F.T. (2016). Switching health insurers: the role of price, quality and consumer information search. European Journal of Health Economics 17(3): 339-353.

Brammli-Greenberg, S., Glazer, J., Waitzberg, R. (2019). Modest risk-sharing significantly reduces health plans' incentives for service distortion. The European Journal of Health Economics 20: 1359-1374.

Breyer, F., Bundorf, K., Pauly, M.V. (2012). Health care spending risk, health insurance, and payment to health plans. In: Pauly, M.V., McGuire, T.G., Barros, P.P. (Eds.), The Handbook of Health Economics, vol. 2 pp 691-762. Elsevier, Waltham.

Brown, J., Duggan, M., Kuziemko, I., Woolston, W. (2014). How does risk selection respond to risk adjustment? New evidence from the medicare advantage program. American Economic Review 104(10), 3335-3364. doi: 10.1257/aer.104.10.3335

Buchner, F., Goepffarth, D., Wasem, J. (2013). The new risk adjustment formula in Germany: implementation and first experiences. Health Policy109:253-262. http://dx.doi.org/10.1016/j.healthpol.2012.12.001

Buchner, F., Wasem, J., Schillo, S. (2017). Regression trees identify relevant interactions: can this improve the predictive performance of risk adjustment? Journal of Health Economics 26, 74-85. doi: 10.1002/hec.3277

Cao, Z., McGuire, T. (2003). Service-level selection by HMOs in Medicare. Journal of Health Economics. 22 (6), 915–931.

Carey, C. (2017a). Technological change and risk adjustment: benefit design incentives in Medicare part D. American Economic Journal: Economic Policy 9 (1), 38–73.

Carey, C. (2017b). Time to Harvest: Evidence on Consumer Choice Frictions from a Payment Revision in Medicare Part D, Working Paper. Retrieved from https://drive.google.com/file/d/0B2TeS7lispKBQ21RLV9PckE2eWs/view.

Decoralis, F., Guglielmo, A. (2017). Insurers' response to selection risk: Evidence from Medicare enrollment reforms. Journal of Health Economics 56, 383-396.

Deming, W.E. (1943). Statistical Adjustment of Data. New York: Wiley.

Duijmelinck, D.M.I.D., Mosca, I., Ven, W.P.M.M. van de. (2015). Switching benefits and costs in competitive health insurance markets: a conceptual framework and empirical evidence from the Netherlands. Health Policy 119:664-671

Eggleston, K., Bir, A. (2009). Measuring selection incentives in managed care: evidence from the Massachusetts state employees insurance program. Journal of Risk and Insurance 76 (1), 159–175.

Eijkenaar, F., van Kleef, R.C., Van Veen, S.H.C.M., van Vliet, R.C.J.A. (2015). Onderzoek risicoverevening 2016: Berekening normbedragen. iBMG/ Erasmus Universiteit Rotterdam, Rotterdam.

Eijkenaar, F., van Vliet, R.C.J.A. (2017). Improving risk equalization using information on physiotherapy diagnoses. European Journal of Health Economics. doi: 10.1007/s10198-017-0874-x

Einav, L., Finkelstein A. (2011). Selection in insurance markets: theory and empirics in pictures. Journal of economic perspectives 25(1), 115-138. doi: 10.1257/jep.25.1.115

Einav, L., Finkelstein, A., Mullainathan, S., Obermeyer, Z. (2018). Predictive modeling of U.S. health care spending in late life. Science 360:1462-1465.

Ellis, R.P., Fiebig, D.G., Johar, M.J., Jones, G., Savage, E. (2013a). Explaining health care expenditure variation: large-sample evidence using linked survey and health administrative data. Health Economics 22, 1093-1110. doi: 10.1002/hec.2916

Ellis, R.E., Jiang, S., Kuo, T. (2013b). Does service level spending show evidence of selection across health plan types? Applied Economics 45:1701-1712. https://doi.org/10.1080/00036846.2011.636023

Ellis, R.P., Martins, B., Zhu, W. (2017). Demand elasticities and service selection incentives among competing private health plans. Journal of Health Economics 56:352-367. https//doi.org/10.1016/j.jhealeco.2017.09.006

Ellis, R.P., Martins, B., Rose, S. (2018). Risk adjustment for health plan payment. In: McGuire, T.G., van Kleef, R.C. (eds.) Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice, pp. 55-104. Elsevier Publishing, Amsterdam.

Ellis, R.P., McGuire, T.G. (2007). Predictability and predictiveness in health care spending. Journal of Health Economics 26:25-48

Enthoven, A.C., van de Ven, W.P.M.M. (2007). Going Dutch – managed-competition health insurance in The Netherlands. The New England Journal of Medicine 357(24), 2421-2423. doi: 10.1056/NEJMp078199

Farid, M.S., McGuire,T.G. (2018). Tail people: the extremely under and overcompensated in individual health insurance markets. In process

Frank, R.G., Glazer, J., McGuire, T.G. (2000). Measuring adverse selection in managed health care. Journal of Health Economics 19 (6), 829–854

Geruso, M., Layton, T. (2015). Upcoding: Evidence from medicare on squishy risk adjustment. NBER, Working Paper 21222

Geruso, M., Layton, T., Prinz, D. (2016). Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges. NBER, Working Paper 22832

Geruso, M., Layton, T., Prinz, D. (2019). Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges. American Economic Journal: Economic Policy 11(2):64-107

Geruso, M, & McGuire, T.G. (2016). Tradeoffs in the design of health plan payment systems: fit, power and balance. Journal of Health Economics 47: 1-19

Glazer, J., McGuire, T.G. (2000). Optimal risk adjustment in markets with adverse selection: an application to managed care. The American Economic Review 90(4), 1055-1071. doi: 10.1016/j.jhealeco.2006.01.005

Glazer, J., McGuire, T.G. (2002). Setting health plan premiums to ensure efficient quality in health care: minimum variance optimal risk adjustment. Journal of public Economics 84,153-173.

Han, T., Lavetti, K. (2017). Does part D abet advantageous selection in medicare advantage? Journal of Health Economics 56, 368-382

Handel, B.R. (2013). Adverse selection and inertia in health insurance markets: when nudging hurts. American Economic Review. 103 (7): 2643-2682. Doi: http://dx.doi.org/10.1257/aer.103.7.2643

Handel, B.R., Kolstad, J.T. (2015). Health insurance for humans: information frictions, plan choice and consumer welfare. American Economic Review 105(8):2449-2500

Handel, B.R., J.T. Kolstad, J. Spinnewijn. (2018). Information frictions and adverse selection: policy interventions in health insurance markets. Review of Economics and Statistics. Doi: https://doi.org/10.1162/rest_a_00773

Heiss, F., McFadden, D., Winter, J., Wupperman, A., Zhou, B. (2016). Inattention and switching costs as sources of inertia in Medicare part D. NBER, Working paper 22765

Izrael, D, Hoaglin, D.C., Battaglia, M.P. (2000). A SAS Macro for Balancing a Weighted Sample. Proceedings of the Twenty-Fifth Annual SAS Users Group International Conference, Cary, NC, SAS Institute Inc., pp. 1350-1355

Kaur, L., McGuire, T.G., Beck, K. (2020). Extreme under and overcompensation in morbidity-based health plan payments: the case of Switzerland. Health policy 124:61-68

Kautter, J., Pope, G.C., Ingber, M., Freeman, S., Patterson, L., Cohen, M., Keenan, P. (2014). The HHS-HHC risk adjustment model for individual and small group markets under the affordable care act. Medicare & Medicaid Research Review 4(3). doi: 10.5600/mmrr.004.03.a03

Ketcham, J.D., Lucarelli, C., Miravete, E.J., Roebuck, M.C. (2012). Sinking, Swimming, or learning to swim in Medicare part D. American Economic Review 102(6): 2639-2673 https://doi.org/10.1257/aer.102.6.2639

Kleef, R.C. van, Eijkenaar, F., van Vliet, R.C.J.A. (2017a). Risicoverevening 2016: uitkomsten op subgroepen uit de gezondheidsmonitor 2012. Rotterdam: iBMG\ Erasumus Universiteit Rotterdam.

Kleef, R.C. van, Eijkenaar, F., van Vliet, R.C.J.A. (2019). Selection incentives for health insurers in the presence of sophisticated risk adjustment. Medical Care Research and Review. https://doi.org/10.1177/1077558719825982

Kleef, R.C. van, Eijkenaar, F., van Vliet, R.C.J.A., van de Ven, W.P.M.M. (2018a). Health plan payment in the Netherlands. In: McGuire, T.G., van Kleef, R.C. (eds.) Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice, pp. 397-430. Elsevier Publishing, Amsterdam

Kleef, R.C. van, McGuire, T.G., van Vliet, R.C.J.A., van de Ven, W.P.M.M. (2017b). Improving risk equalization with constrained regression. European Journal of Health Economics 18:1137-1156. DOI: 10.1007/s10198-016-0859-1

Kleef, R.C. van, van de Ven, W.P.M.M., van Vliet, R.C.J.A. (2013a). 'Risk selection in a regulated health insurance market: a review of the concept, possibilities and effects.' Expert Reviews Pharmacoeconomics Outcomes Research 13(6):743-752.

Kleef, R.C. van, van Vliet, R.C.J.A., Eijkenaar, F., Nielen, M.M.J. (2018b). Gebruik van diagnose-informatie uit huisartsenregistraties in de risicoverevening via constrained regression (Dutch). ESHPM, Rotterdam

Kleef, R.C. van, van Vliet, R.C.J.A., van de Ven, W.P.M.M. (2012). Risicoverevening tussen zorgverzekeraars: kwantificering modelverbeteringen 1993-2011. Tijdschrift voor Gezondheidswetenschappen 5, 311-326. doi: 10.1007/s12508-012-0110-0

Kleef, R.C. van, van Vliet, R.C.J.A., van de Ven, W.P.M.M. (2013b). Risk selection in the Netherlands: an empirical evaluation. Expert Reviews Pharmacoeconomics Outcomes Research 13(6):829-839.

Kleef, R.C. van, van Vliet, R.C.J.A., van de Ven, W.P.M.M. (2014). Risicoverevening 2014 voor somatische zorg: Analyse van uitkomsten op subgroepniveau. Rotterdam: iBMG\ Erasumus Universiteit Rotterdam.

Kleef, R.C. van, van Vliet, R.C.J.A. (2022). How to deal with persistently low/high spenders in health plan payment. Health Economics. https://doi.org/10.1002/hec.4477

Kling, J.R., Mullainathan, S., Shafir, E., Vermeulen, L.C., Wrobel, M.V. (2012) Comparison friction: experimental evidence from Medicare drug plans. Quarterly Journal of Economics 127(1): 199-235 doi:10.1093/qje/qjr055.

KPMG. (2014) Evaluatie Zorgverzekeringswet. KPMG, Amstelveen

KPMG. (2020) Onderzoek restprobleem risicoverevening fase 1. KPMG, Amstelveen

Lamers, L.M. (1999). Risk-adjusted capitation based on the diagnostic cost group model: an empirical evaluation with health survey information. Health services research 33(6), 1727-1744.

Laske-Aldershof, T., Schut, E., Beck, K., Greß, S., Shmueli, A., Voorde, C. van de. (2004). Consumer mobility in social health insurance markets. A five country comparison. Applied health econ health policy 3(4):229-241

Lavetti, K., Simon, K. (2016). Strategic formulary design in Medicare Part D plans. Natl. Bureau Econ. Res., Working Paper 22338

Layton, T.J., Ellis, R.P., McGuire, T.G., van Kleef, R.C. (2017). Measuring efficiency of health plan payment systems in managed competition health insurance markets. Journal of Health Economics 56, 237-255.

Layton, T.J., McGuire, T.G., van Kleef, R.C.: (2018). Deriving risk adjustment payment weights to maximize efficiency of health insurance markets. Journal of Health Economics 61, 93-110.

Lissenden, B. (2019). The effect of cancer diagnosis on switching health insurance in medicare. Health economics 28, 339-349.

McGuire, T.G., Glazer, J., Newhouse, J.P., Normand, S., Shi, J., Sinaiko, A.D., Zuvekas, S.H. (2013): Integrating risk adjustment and enrollee premiums in health plan payment. Journal of Health Economics 32, 12363-1277.

McGuire, T.G., Kleef, R.C. van. (2018a). Risk adjustment, risk sharing and premium regulation in health insurance markets. Theory and practice. Elsevier Publishing, Amsterdam.

McGuire, T.G., Kleef, R.C. van. (2018b). Regulated competition in health insurance markets: Paradigms and ongoing issues. In: McGuire, T.G., van Kleef, R.C. (eds.) Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice, pp. 397-430. Elsevier Publishing, Amsterdam

McGuire, T.G., Kleef, R.C. van. (2018c). Risk sharing. In: McGuire, T.G., Kleef, R.C. van (eds.), Risk adjustment, risk sharing and premium regulation in health insurance markets. Theory and practice pp.105-132. Elsevier, Amsterdam.

McGuire, T.G., Newhouse, J.P., Normand, S., Shi, J., Zuvekas, S. (2014). Assessing incentives for service level selection in private health insurance exchanges. Journal of Health Economics 35: 47-63. Doi: 10.1016/j.jhealeco.2014.01.009

McGuire, T.G., Schillo, S., van Kleef, R.C. (2020). Reinsurance, repayments and risk adjustment in individual health insurance. American Journal of Health Economics. https://doi.org/10.1086/706796

McWilliams, J.M., Hsu, J., Newhouse, J.P. (2012). New risk-adjustment system was associated with reduced favorable selection in medicare advantage. Health affairs 31(12), 2630-2640. doi: 10.1377/hlthaff.2011.1344

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

Newhouse, J.P., Price, M., Huang, J., McWilliams, M., Hsu, J. (2012). Steps to reduce favorable risk selection in medicare advantage largely succeeded, boding well for health insurance exchanges. Health affairs 31(12), 2618-2628. doi:10.1377/hlthaff.2012.0345

Newhouse, J.P., Price, M., McWilliams, J.M., Hsu, J., McGuire, T.G. (2015). How much favorable selection is left in medicare advantage? American Journal of Health Economics 1(1), 1-26. doi: 10.1162/ajhe_a_00001

Pauly, M.V. (1998) Managed care makertpower and monopsony.' Health Services Research 33 (5):137-147 (1998)

Polder, J.J., Barendregt, J.J., Van Oers, H. (2006). Health care costs in the last year of life- The Dutch experience. Social Science & Medicine 63:1720-1731.

Public Health Monitor 2012 of the Community Health Services, Statistics Netherlands and the National Institute for Public Health and the Environment

Rose, S. (2016). A machine learning framework for plan payment risk adjustement. Health Services Research 51 (6): 2358-2374.

Rothschild, M., Stiglitz, J. (1976). Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. The Quarterly Journal of Economics 90 (4), 629–649

Schillo, S., Lux, G., Wasem, J., Buchner, F. (2016). High cost pool or high cost groups - how to handle high(est) cost cases in a risk adjustment mechanism? Health Policy 120: 141-147.

Schors, W. van der, Brabers, A.E.M., de Jong, J.D. (2020). Does the chronically ill population in the Netherlands switch their health insurers as often as the general population? Emperical evidence from a nationwide survey study. BMC Health Services Research 20 (376) Doi: 10.1186/s12913-020-05228-z

Schut, E., M. Varkevisser, M. (2009). Marktordening in de gezondheidszorg. In: Schut, E., Rutten, F. (eds.) Economie van de gezondheidszorg, pp. 247-288. Elsevier gezondheidszorg, Maarssen.

Shepard, M. (2016). Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange, NBER Working Paper, number 22600

Shmueli, A., Messika, D., Zmora, I., Oberman, B. (2010). Health care costs during the last 12 months of life in Israel: estimation and implications for risk adjustment. International Journal of Health Care Finance and Economics 10:257-273. DOI: 10.1007//s10754-010-9080-4

Shmueli, A., Nissan-Engelcin, E. (2013). Local availability of physicians' services as a tool for implicit risk selection. Social Science & Medicine 84, 53–60

Speybroeck, N. (2012). Classification and regression trees. International Journal of Public Health 57, 243-246. doi: 10.1007/s00038-011-0315-z

Stam, P.J.A., van Vliet, R.C.J.A., van de Ven, W.P.M.M. (2010). Diagnostic, pharmacy-based and self-reported health measures in risk equalization models. Medical care 48 (5), 448-457. doi: 10.1097/MLR.0b013e3181d559b4

Stam. P., Ismail, I., Visser, J., Portrait, F., Koolman, X. (2021). Machine learning vereist beter risicoverevening zorgverzekeraars. Economisch Statistische Berichten (ESB) 106(4974).

Stooker, T., van Acht, J.W., van Barneveld, E.M., van Vliet, R.C.J.A., Hessing, D.J., Bussenbach, J.J.V. (2001). Costs in the last year of life in the Netherlands. Inquiry 38(1):73-80. doi: 10.5034/inquiryjrnl_38.1.73.

Veen, S.H.C.M. van, van Kleef, R.C., van de Ven, W.P.M.M. (2015a). Improving the prediction model used in risk-equalization: cost and diagnostic information from multiple prior years. Eur J Health Econ 16, 201–218. https://doi.org/10.1007/s10198-014-0567-7

Veen, S.H.C.M. van, van Kleef, R.C., van de Ven, W.P.M.M., van Vliet, R.C.J.A. (2015b). 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 72(2):220-243.

Veen, S.H.C.M. van, van Kleef, R.C., van de Ven, W.P.M.M., van Vliet, R.C.J.A. (2016). Residual expenses from multiple prior years. In: Veen, S.H.C.M. (eds) Evaluating and improving the predictive performance of risk equalization models in health insurance markets, pp. 185-224. Optima grafische communicatie, Rotterdam.

Ven, W.P.M.M. van de (2011). Risk adjustment and risk equalization: what needs to be done? Health Economics, Policy and Law 6, 147-156. doi:10.1017/S1744133110000319

Ven, W.P.M.M. van de, Beck, K., Buchner, F., Chernichovsky, D., Gardiol, L., Holly, A., Lamers, L.M. Schokkaert, E., Shmueli, A., Spycher, S., van de Voorde, C., van Vliet, R.C.J.A., Wasem, J., Zmora, I. (2003). Risk equalization and risk selection on the sickness fund insurance market in five European countries. Health Policy 65:75-98.

Ven, W.P.M.M. van, Beck, K., Buchner, F., Schokkaert, E., Schut, F.T., Shmueli, A., Wasem, J. (2013). Preconditions for efficiency and affordability in competitive healthcare markets: Are they fullfilled in Belgium, Germany, Israel, the Netherlands and Switzerland? Health Policy 109, 226-245. doi: 0.1016/j.healthpol.2013.01.002

Ven W.P.M.M. van de, Beck, K., van de Voorde, C., Wasem, J., Zmora, I.: (2007). Risk adjustment and risk selection in Europe 6 years later. Health Policy 83:162-179. DOI: 10.1016/j.healthpol.2006.12.004

Ven, W.P.M.M. van de, Ellis, R.P. (2000). Risk adjustment in competitive health insurance markets. In: Culyer, A.J., Newhouse, J.P., (Eds.), Handbook of Health Economics, 755-845. Elsevier, Amsterdam.

Ven, W.P.M.M. van de, Hamstra, G., Kleef, R.C. van, Reuser, M. Stam, P. (2021). The goal of risk equalization in regulated competitive health insurance markets, *forthcoming.*

Ven, W.P.M.M. van de, van Kleef, R.C. (2016). De rol van zorgverzekeringen. In: Schut, F.T., Varkevisser, M. (eds.) Economie van de gezondheidszorg. Springer, Houten p. 181-234.

Ven, W.P.M.M. van de, van Kleef, R.C., van Vliet, R.C.J.A. (2015). Risk selection threatens quality of care for certain patients: lessons from Europe's health insurance exchanges. Health Affairs 34, 1713–1720.

Ven, W.P.M.M. van de, Schut, F.T. (2008). Universal mandatory health insurance in the Netherlands: a model for the United States? Health Affairs 27 (3):771-781

Ven, W.P.M.M. van de, Schut, F.T. (2011). Guaranteed access to affordable coverage in individual health insurance markets'. In: Glied, S., Smith, P.C. (eds.) The Oxford handbook of health economics, pp. 380-404. Oxford University Press, Oxford.

Ven, W.P.M.M. van de, van Vliet, R.C.J.A., van Kleef, R.C. (2017). How can the regulator show evidence of (no) risk selection in health insurance markets? Conceptual framework and empirical evidence. European Journal of Health Economics 18, 167-180. doi: 10.1007/s10198-016-0764-7

Vliet, R.C.J.A. van. (2006). Free choice of health plan combined with risk-adjusted capitation payments: are switchers and new enrollees good risks? Health economics 15:763-774

Vliet, R.C.J.A. van, Lamers, L.M. (1998). The high cost of death: should health plans get higher payments when members die. Medical Care 36(10):1451-1460.

Volksgezondheidenzorg.info: Health monitor adults, public health services, Statistics Netherlands, and National Institute for Public health and the Environment. Volksgezondheidenzorg.info. https://bronnen.zorggegevens.nl/Bron?naam=Gezondheidsmonitor-Volwassenen%2C-GGD%E2%80%99en%2C-CBS-en-RIVM. Accessed March 12, 2018.

Withagen-Koster, A.A., van Kleef, R.C., Eijkenaar, F. (2018). Examining unpriced risk heterogeneity in the Dutch health insurance market. European Journal of Health Economics. https://doi.org/10.1007/s10198-018-0979-x

Withagen-Koster, A.A., van Kleef, R.C., Eijkenaar, F. (2020). Incorporating self-reported health measures in risk equalization through constrained regression. European Journal of Health Economics. https://doi.org/10.1007/s10198-019-01146-y


S

Summary

Many countries have reformed their health insurance systems towards a system of regulated competition. Within this system health insurers compete on price (premium) and on the composition and quality of the contracted provider network, while regulation safeguards public goals such as affordability and accessibility of basic coverage. Given community-rated premiums, risk equalization is an important regulatory tool to mitigate risk-selection incentives. However, even under the most sophisticated risk equalization models currently used around the world, there are still predictable profits and losses present on healthy and chronically ill individuals, respectively.

Chapter 1 introduces the topic and explains why it is important to (further) mitigate risk-selection incentives. In this dissertation we use a large health survey to evaluate and examine possible improvements for sophisticated risk equalization. This large health survey provides a unique opportunity to do so. The reason is that evaluating and examining possible improvements for risk equalization requires data not explicitly used as a basis for risk adjusters. Health survey information typically is not included in risk equalization and can therefore be used to do so. This chapter also introduces the research questions addressed in this dissertation.

Chapter 2 starts by examining to what extent the health survey information can explain variation in residual spending. Specifically, we research to what extent unpriced risk heterogeneity is present, which refers to predictable variation in health care spending not reflected in either premiums by insurers, or risk equalization payments. The survey information allows for explaining and predicting residual spending of the risk equalization model. The analyses yield three main findings: (1) the health survey information is able to explain some residual spending of the risk equalization model, (2) unpriced risk heterogeneity exists both in morbidity and in non-morbidity groups (as defined by the risk equalization model), and (3) unpriced risk heterogeneity increases with predicted spending by the risk equalization model. These findings imply that the sophisticated Dutch risk equalization model does not completely remove unpriced risk heterogeneity.

In chapter 3 we research to what extent predictable profits and losses for selective groups exist after sophisticated risk equalization from a multiple contract period perspective. Typically, empirical studies evaluating selection incentives have quantified group-level predictable profits and losses for one contract period. However, due to switching barriers, a multiple contract period perspective might be more relevant. We find that most of the groups of chronically ill are persistently unprofitable on average, while the healthy group is on average persistently profitable. This implies

that selection incentives might be stronger than initially thought, underscoring the necessity of reducing predictable profits and losses for the adequate functioning of competitive social health insurance markets.

Chapter 4 focusses on a specific possible driver of remaining predictable profits and losses: end-of-life spending. More specifically, we examine the extent to which end-of-life spending contributes to predictable profits and losses for selective groups identifiable in the health survey. We do so by simulating the predictable profits/losses for these groups *with* and *without* end-of-life spending while correcting for the overall spending difference between these two situations. Our main finding is that – even under a sophisticated risk-equalization model – end-of-life spending contributes to predictable losses for specific chronic conditions.

In chapter 5 and 6 we turn our attention to possible improvements of sophisticated risk equalization using health survey information. Self-reported health measures are often considered not suitable to include as explicit risk adjusters in risk-equalization models. Therefore, in chapter 5 we research an alternative way to exploit this information, namely through 'constrained regression' (CR). We estimate five CR models and compare these models with the actual Dutch risk equalization model of 2016 estimated by ordinary least squares (OLS). In the CR-models the estimated coefficients are restricted such that the under-/overcompensation for groups based on self-reported general health is reduced by 20, 40, 60, 80 or 100 percent. Our results show that CR can improve outcomes for groups that are not explicitly flagged by risk adjuster variables but worsens outcomes for groups that are explicitly flagged by risk adjuster variables. Using a new standardized metric that summarizes predictable profits/losses for both types of groups, we find that CR models using the lighter constraints can lead to better outcomes than OLS.

In chapter 6 we research a different strategy to exploit health survey information in risk equalization, namely high-risk pooling. The essence of our high-risk pooling concept is that insurers can ex-ante assign predictably unprofitable individuals to a pool using information from a health survey. We evaluate the effect of five alternative pool sizes based on predicted residual spending post risk equalization on insurers' incentives for risk selection and cost control and compare them to the situation without high-risk pooling. Our main conclusion is that high-risk pooling has the potential to considerably reduce remaining risk selection incentives at the expense of a relatively modest reduction of incentives for cost control.

Chapter 7 answers the research questions by summarizing the main findings and conclusions of this dissertation. Chapter 8 then discusses the main findings, presents implications for policy and practice and provides some suggestions for further research. We emphasize the importance of regulators to explicitly weigh the welfare effects of the various methods of risk equalization. The reason is that innovative strategies to (further) mitigate risk selection incentives present the regulator with tradeoffs. For instance, the regulator should weigh the welfare effects of reducing incentives for risk selection and cost control. To refine and evaluate such strategies, explicit normative choices in risk equalization must be made.


S

Samenvatting

Een aantal landen in de wereld hebben een systeem van gereguleerde concurrentie geïmplementeerd in hun zorgverzekeringsmarkt. In dit systeem concurreren zorgverzekeraars op prijs (premie) en de samenstelling en kwaliteit van het gecontracteerde netwerk van zorgaanbieders binnen wettelijke kaders die publieke belangen beschermen, zoals betaalbaarheid en toegankelijkheid van zorg. Gegeven een doorsneepremie per zorgpolis, is risicoverevening belangrijk om prikkels tot risicoselectie tegen te gaan. Echter, zelfs onder de meest geavanceerde risicovereveningsmodellen bestaan er nog voorspelbare winsten en verliezen op respectievelijk gezonde en ongezonde groepen.

Hoofdstuk 1 introduceert het onderwerp en legt uit waarom het belangrijk is om risicoselectie prikkels verder te verminderen. In dit proefschrift gebruiken we data van een grote gezondheidsenquête om geavanceerde risicoverevening te evalueren en mogelijke verbeteringen te onderzoeken. Deze grote gezondheidsenquête geeft een unieke mogelijkheid om dit te doen aangezien dit soort informatie over het algemeen niet gebruikt wordt in de risicoverevening. Dit hoofdstuk introduceert ook de onderzoeksvragen die in dit proefschrift worden beantwoord.

Hoofdstuk 2 start met het onderzoeken van in hoeverre de informatie uit de gezondheidsenquête variatie in residuele kosten (na risicoverevening) kan verklaren. We hebben hierbij specifiek onderzocht in hoeverre risicoheterogeniteit aanwezig is die niet tot uitdrukking komt in de premie of de risicovereveningsbijdrage. De informatie uit de gezondheidsenquête hebben we gebruikt om de residuele kosten van het risicovereveningsmodel te verklaren en te voorspellen. De drie belangrijkste bevindingen van onze analyses in dit hoofdstuk zijn dat (1) de informatie uit de gezondheidsenquête in staat is om nog iets van de residuele kosten te verklaren, (2) risicoheterogeniteit aanwezig is in zowel de morbiditeitsgroep als de niet-morbiditeitsgroep (zoals gedefinieerd door het risicovereveningsmodel), en dat (3) risicoheterogeniteit toeneemt met de voorspelde kosten volgens het risicovereveningsmodel. Deze bevindingen impliceren dat het geavanceerde Nederlandse risicovereveningsmodel niet alle risicoheterogeniteit wegneemt.

In hoofdstuk 3 onderzoeken we in hoeverre er voorspelbare winsten en verliezen bestaan voor selectieve groepen na geavanceerde risicoverevening over meerdere contractperioden. Over het algemeen evalueren empirische studies risicoselectie prikkels door de voorspelbare winsten en verliezen voor groepen voor één contractperiode in kaart te brengen. Echter, vanwege (gepercipieerde) barrières voor verzekerden om over te stappen naar een andere zorgverzekering, is een perspectief dat voorspelbare winsten en verliezen over meerdere contractperioden evalueert

mogelijk relevanter. Onze bevindingen laten zien dat de meeste groepen chronisch zieken gemiddeld genomen over meerdere contractperioden worden ondergecompenseerd, terwijl gezonde groepen gemiddeld genomen over meerdere contractperioden worden overgecompenseerd. Dit impliceert dat risicoselectie prikkels mogelijk sterker zijn dan initieel gedacht, wat de noodzaak voor het verminderen van voorspelbare winsten en verliezen benadrukt voor het adequaat functioneren van competitieve sociale zorgverzekeringsmarkten.

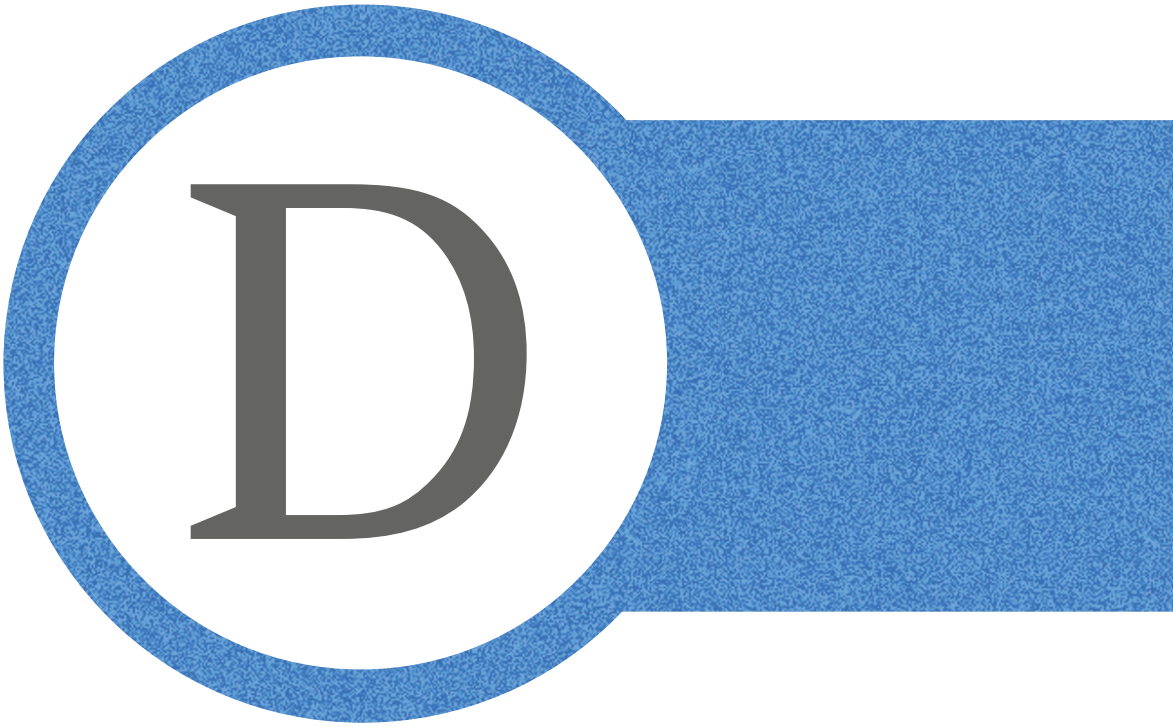
Hoofdstuk 4 focust op een specifieke mogelijke driver van voorspelbare winsten en verliezen: kosten in de laatste levensfase. We onderzoeken in hoeverre kosten in de laatste levensfase bijdragen aan voorspelbare winsten en verliezen voor selectieve groepen die we identificeren in de gezondheidsenquête. Dit hebben we gedaan door de voorspelbare winsten/verliezen te simuleren *met* en *zonder* kosten in de laatste levensfase, waarbij we corrigeren voor het verschil in kostenniveau tussen de twee situaties. Onze belangrijkste bevinding in dit hoofdstuk is dat – zelfs onder geavanceerde risicoverevening – kosten in de laatste levensfase bijdragen aan voorspelbare verliezen voor specifieke chronische aandoeningen.

In hoofdstuk 5 en 6 richten we onze aandacht op mogelijke verbeteringen van geavanceerde risicoverevening op basis van de informatie uit de gezondheidsenquête. Omdat zelf-gerapporteerde gezondheidsmaten over het algemeen niet geschikt zijn om als risicovereveningkenmerk in het model op te nemen, onderzoeken we in hoofdstuk 5 daarom een alternatieve manier om informatie uit een gezondheidsenquête te includeren in risicoverevening, namelijk via constrained regression (CR). We hebben vijf CR modellen geschat en de uitkomsten daarvan vergeleken met het daadwerkelijke Nederlandse risicovereveningsmodel van 2016, dat geschat wordt met de kleinste kwadraten methoden (ordinary least squares, OLS). Er wordt een dusdanige restrictie op de geschatte coëfficiënten van de CR-modellen gelegd dat de onder-/overcompensatie voor groepen gebaseerd op zelf-gerapporteerde algemene gezondheid wordt verminderd met 20, 40, 60, 80 en 100 procent. De resultaten in dit hoofdstuk laten zien dat CR de uitkomsten kan verbeteren voor groepen die niet expliciet opgenomen zijn in het risicovereveningsmodel, maar ook dat de uitkomsten voor groepen die wel expliciet in het model zijn opgenomen verslechteren. Op basis van een nieuwe, gestandaardiseerde maat die de voorspelbare winsten/verliezen voor beide type groepen sommeert, concluderen we echter dat lichte restricties tot betere uitkomsten kunnen leiden dan OLS.

Hoofdstuk 6 onderzoekt een andere strategie om informatie uit de gezondheidsenquête te includeren in risicoverevening, namelijk door een vorm van hoge risico

verevening (HRV). De kern van ons HRV-concept is dat zorgverzekeraars ex ante voorspelbaar verliesgevende verzekerden aan een pool kunnen toewijzen, gebruikmakend van informatie uit een gezondheidsenquête. We hebben de effecten van vijf HRV-varianten op de prikkels voor verzekeraars voor risicoselectie en kostenbeheersing gesimuleerd, en deze vergeleken met de situatie zonder HRV. Onze belangrijkste bevinding in dit hoofdstuk is dat HRV de potentie heeft om een aanzienlijk deel van de prikkels voor risicoselectie te verminderen, ten koste van een relatief kleine afname van de prikkels voor kostenbeheersing.

Hoofdstuk 7 beantwoordt elk van de onderzoeksvragen die zijn gesteld en vat de belangrijkste bevindingen en conclusies samen. Hoofdstuk 8 bediscussieert vervolgens de belangrijkste bevindingen, presenteert implicaties voor beleid en praktijk en geeft suggesties voor verder onderzoek. Dit hoofdstuk benadrukt tevens het belang voor de overheid om de welvaartseffecten af te wegen van de verschillende manieren om risicoverevening vorm te geven. De reden is dat innovatieve methoden om risicoselectie prikkels te verminderen de overheid met afwegingen confronteert, zoals een afweging tussen de reductie van prikkels voor risicoselectie en doelmatigheid. Normatieve keuzes omtrent deze afwegingen moeten worden gemaakt om innovatieve methoden verder te verfijnen en te evalueren.


D

Dankwoord

Het schrijven van een proefschrift is geen klein klusje. Ondanks dat mijn naam op de omslag staat, heb ik dit werk niet kunnen doen zonder de hulp en steun van een aantal mensen. Bovenal wil ik God danken en prijzen voor zijn zegening en leiding in dit hele traject, welke onmisbaar zijn geweest.

Het schrijven van mijn proefschrift had ik niet kunnen doen zonder de waardevolle begeleiding en feedback van Richard, Frank en Erik. Ik wil jullie hier speciaal bedanken voor de fijne gesprekken die we over de jaren gehad hebben, zowel over de inhoud als over alles wat daarnaast speelde. Ondanks de drukke agenda's maakten jullie altijd tijd om mijn stukken grondig door te nemen en van feedback te voorzien. Ik heb het over de jaren vaker gezegd: ik heb het getroffen met jullie als begeleiders! Daarnaast wil ik René en Wynand ook bedanken voor hun waardevolle feedback op mijn papers. Verder wil ik het ministerie van Volksgezondheid, Welzijn en Sport, Zorgverzekeraars Nederland en de GGD GHOR Nederland bedanken voor het gebruik van de data. Ook wil ik de leden van de beoordelings- en promotiecommissie graag bedanken voor de tijd en moeite die ze hebben genomen om kritisch naar mijn proefschrift te kijken en om te opponeren tijdens mijn verdediging.

Lieve Sanne, wat hebben wij een lol gehad de afgelopen jaren! Ik heb genoten van de koffies en lunches en de goede gesprekken die we hebben gehad. Je hebt een heerlijke nuchtere persoonlijkheid en houding waar ik erg van kan genieten. Je bent een fijne kamergenoot, collega en vriendin!

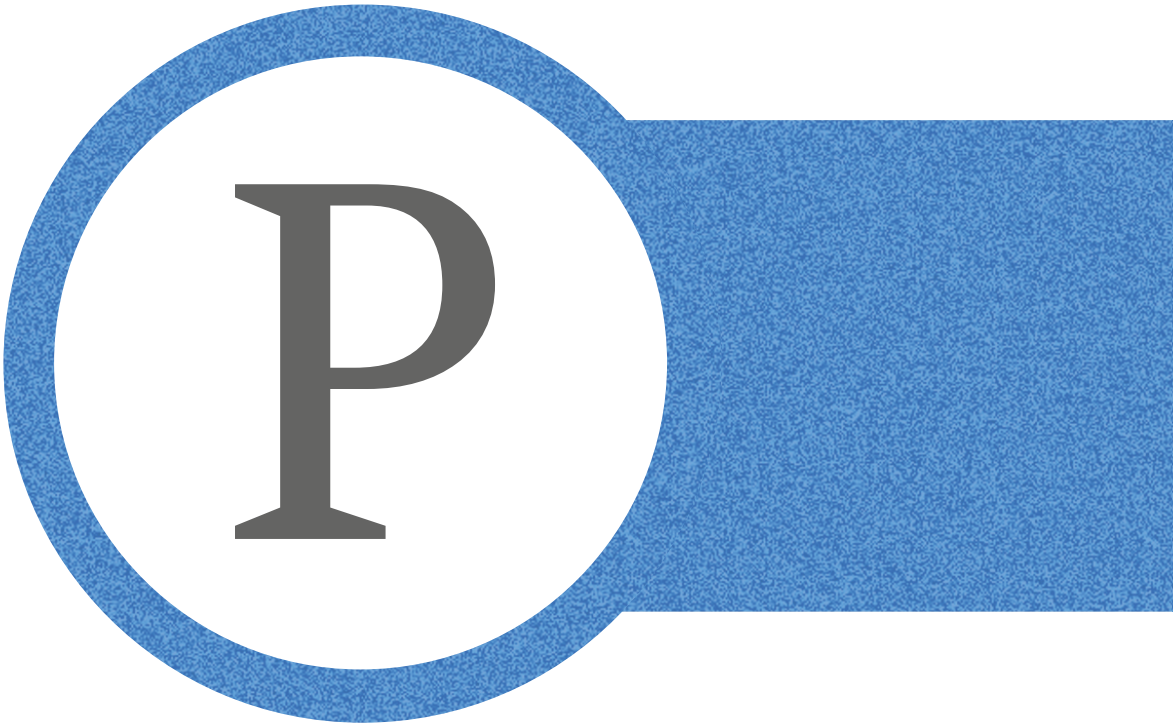
Danielle, ik weet nog goed de eerste dag dat ik binnenkwam bij HSI en jij mij zo hartelijk welkom hebt geheten! Sinds die eerste dag heb je me regelmatig van advies voorzien, waarvoor ik je nog steeds dankbaar ben. Ik heb genoten van de gesprekken die we gehad hebben, van onderwijs tot babysokjes!

Graag wil ik ook alle collega's waar ik mee heb mogen samenwerken bij ESHPM en in het bijzonder van HSI bedanken voor de fijne samenwerking van de afgelopen jaren. In het bijzonder wil ik Newel en Judith bedanken voor de leuke en fijne gesprekken. Marjolein, ik wil jou bedanken voor je advies en je heerlijke down-to-earth houding. Dit heeft me vaker geholpen de promotiewerkzaamheden weer in het juiste perspectief te krijgen dan dat ik tegen je gezegd heb.

Arie, zoals jij is er geen een. Ik ben blij met jou als broer en hoe je mij op je eigen manier hebt gestimuleerd me te blijven ontwikkelen.

Lieve pa en ma, jullie hebben mij van jongs af aan gestimuleerd om mijn best te doen en om mijn grenzen te blijven verleggen. Ik kan wel zeggen dat dit proefschrift hier het resultaat van is. Dank jullie wel voor jullie onvoorwaardelijke liefde en betrokkenheid.

Lieve Rik, dank je wel voor wie je bent. Dank je wel dat je mij onvoorwaardelijk steunt in alles wat ik onderneem en dat je altijd naast me staat. Zonder jouw aanmoedigingen en steun had ik dit proefschrift niet kunnen schrijven.



Portfolio

Education	
Bachelor of Science in 'Beleid en management van de gezondheidszorg', Erasmus University Rotterdam, The Netherlands	2015
Master of Science in 'Health Economics, Policy and Law', specialization 'Health Economics', Erasmus University Rotterdam, The Netherlands	2016

International peer-reviewed publications	
Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. (2018). Examining unpriced risk heterogeneity in the Dutch health insurance market. European Journal of Health Economics. https://doi.org/10.1007/s10198-018-0979-x	2018
Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. (2020). Incorporating self-reported health measures in risk equalization through constrained regression. European Journal of Health Economics. https://doi.org/10.1007/s10198-019-01146-y	2020
Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. (2022). Selection incentives in the Dutch basic health insurance: to what extent does end-of-life spending contribute to predictable profits and losses for selective groups? Medical Care Research and Review, accepted for publication.	2022
Withagen-Koster, A.A., van Kleef, R.C., Eijkenaar, F. Predictable profits and losses in a health insurance market with risk equalization: a multiple-contract period perspective. Submitted	
Withagen-Koster, A.A., Van Kleef, R.C., Eijkenaar, F. An innovative method to mitigate risk selection incentives in health insurance markets with sophisticated risk equalization: high-risk pooling based on health survey information. Submitted	

Contributions to Dutch policy research projects	
Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2017). "Onderzoek risicoverevening 2018: Berekening Normbedragen" Rapport iBMG, Rotterdam: Erasmus Universiteit.	2017
Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2017). "Onderzoek risicoverevening 2018: Overall Toets" Rapport iBMG, Rotterdam: Erasmus Universiteit.	2017
Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2017). "Onderzoek risicoverevening 2018: Gegevensfase" Rapport iBMG, Rotterdam: Erasmus Universiteit.	2017
Cattel, D., Eijkenaar F, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2017). "Onderzoek risicoverevening 2018: Pre-Overall Toets" Rapport iBMG, Rotterdam: Erasmus Universiteit.	2017
Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2018). "Onderzoek risicoverevening 2019: Berekening Normbedragen" Rapport ESHPM, Rotterdam: Erasmus Universiteit.	2018
Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2018). "Onderzoek risicoverevening 2019: Overall Toets" Rapport ESHPM, Rotterdam: Erasmus Universiteit.	2018
Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2018). "Onderzoek risicoverevening 2019: Gegevensfase" Rapport ESHPM, Rotterdam: Erasmus Universiteit.	2018

Eijkenaar F, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2018). "Onderzoek risicoverevening 2019: Pre-Overall Toets" Rapport ESHPM, Rotterdam: Erasmus Universiteit	2018
Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2019). "Onderzoek risicoverevening 2020: Berekening Normbedragen" Rapport ESHPM, Rotterdam: Erasmus Universiteit.	2019
Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2019). "Onderzoek risicoverevening 2020: Overall Toets" Rapport ESHPM, Rotterdam: Erasmus Universiteit.	2019
Cattel, D., F. Eijkenaar, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2019). "Onderzoek risicoverevening 2020: Gegevensfase" Rapport ESHPM, Rotterdam: Erasmus Universiteit.	2019
Kleef, R.C. van, M. Oskam, L.A. Visser, R.C.J.A. van Vliet, A.A. Withagen-Koster (2020). "Grootonderhoud DKG's t.b.v. somatisch model 2021" Rapport ESHPM, Rotterdam: Erasmus Universiteit.	2020
Cattel, D., F. Eijkenaar, M. Oskam, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2020). "Onderzoek risicoverevening 2021: Berekening Normbedragen" Rapport ESHPM, Rotterdam: Erasmus Universiteit.	2020
Cattel, D., F. Eijkenaar, M. Oskam, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2020). "Onderzoek risicoverevening 2021: Overall Toets" Rapport ESHPM, Rotterdam: Erasmus Universiteit.	2020
Cattel, D., F. Eijkenaar, M. Oskam, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2020). "Onderzoek risicoverevening 2021: Gegevensfase" Rapport ESHPM, Rotterdam: Erasmus Universiteit.	2020
Cattel, D., F. Eijkenaar, M. Oskam, R.C. van Kleef, R.C.J.A. van Vliet en A.A. Withagen-Koster (2020). "Onderzoek risicoverevening 2021: pre Overall Toets" Rapport ESHPM, Rotterdam: Erasmus Universiteit. 2020	2020

Projects	
Member of the team that calculates the risk adjustment payment weights for the Dutch Ministry of Health	2017 - 2021
Research for the Dutch Ministry of Health: Maintenance of the DCG risk adjuster	2020

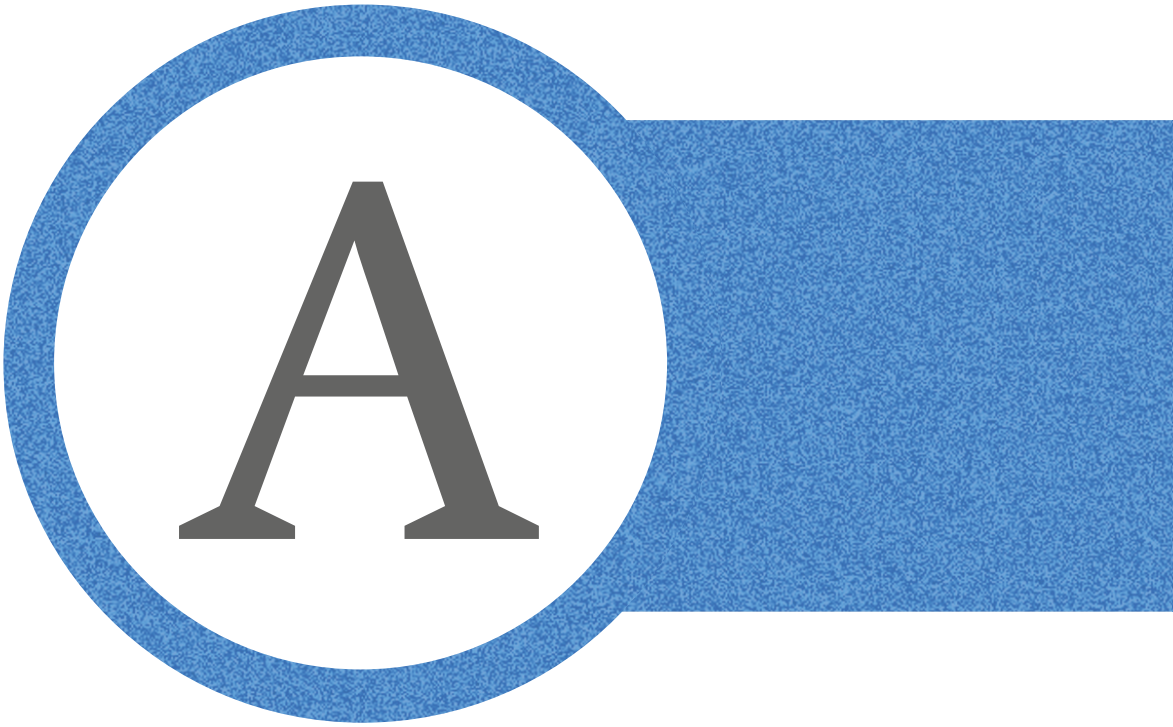
PhD Training	
Training course: A successful doctoral track ("Ready in 4 years"), Erasmus University Rotterdam	2016
SAS base programming fast-track, SAS institute Huizen	2016
Academic writing in English, Erasmus University Rotterdam	2017
Self-presentation: Presenting yourself and your research, Erasmus University Rotterdam	2017
Basiccourse didactics, Erasmus University Rotterdam	2017
Bachelorthesis supervision (intervision), Erasmus University Rotterdam	2017
Bachelorthesis supervision (intervision), Erasmus University Rotterdam	2018
Workshop 'Personal leadership', Erasmus University Rotterdam	2018
Workshop '150 word pitch', Erasmus University Rotterdam	2018
Group dynamics, Erasmus University Rotterdam	2018

MOOC econometrics: methods and applications, Erasmus University Rotterdam	2019
Microlab: Storytelling in higher education, Erasmus University Rotterdam	2019
Workshop 'Make it easier for yourself', Erasmus University Rotterdam	2019
Workshop 'negotiating', Erasmus University Rotterdam	2019
Analytic storytelling, Erasmus University Rotterdam	2019

Teaching	
Tutor in the bachelor cource 'Multivariate analyse'	2016 - 2018
Quantitative research rapport (KLO) pre-master level: Supervision of students	2016 - 2018
Quantitative research rapport (KLO) pre-master level: Course coordinator	2019 - 2021
Bachelor thesis supervision	2016 - 2021
Tutor in the bachelor course 'Blok 1 de Nederlandse gezondheidszorg'	2018 - 2020

Conferences, seminars and presentations	
LoLa, Rotterdam	2017
Dutch Ministery of Health, Welfare and Sports, The Hague	2017
Risk Adjustment Network, The Hague, Netherlands (presentation)	2017
Risk Adjustment Network, Tel Aviv, Israel (presentation)	2018
Risk Adjustment Network, Maine, United States of America (presentation)	2019
Risk Adjustment Network, online (presentation)	2020
Yearly seminar 'Uitvoering risicoverevening' (ZIN)	2017 - 2018
Vereniging voor Gezondheidseconomie, Rotterdam	2018
Vereniging voor Gezondheidseconomie, Zeist	2019
Vereniging voor Gezondheidseconomie, Utrecht	2020

Other	
Organization of the post academic course risk equalization	2018 - 2021
Maintaining the information on the website for the section HSI	2017 - 2021
Member of the strategic workgroup 'future-oriented education'	2019
Organisation of EuHEa PhD conference	2020 - 2021
Career coaching, Erasmus University Rotterdam	2021


A

About the author

Anja Withagen-Koster obtained her master's degree in Health Economics, Policy and Law (specialization health economics) in 2016. That same year she started as a PhD student at Erasmus School of Health Policy & Management (ESHPM). During her PhD trajectory she evaluated and studied possible improvement of the sophisticated Dutch risk equalization model using a large and rich health survey. The results of these studies are published in several international peer-reviewed journals. Anja has also presented the outcomes of these studies at several international conferences. In addition, she taught various bachelor and pre-master courses and supervised multiple bachelor theses. Currently, Anja is working as a policy adviser at the Dutch Ministry of Health, Welfare and Sports, specifically focusing on risk equalization.