Evaluating and Mitigating Incentives for Risk Selection in Health Insurance Markets with Risk Equalization

Het evalueren en verminderen van selectieprikkels in zorgverzekeringsmarkten met risicoverevening

Michel Oskam

ISBN 978-94-6534-597-0

Lay-out and printing by Optima Grafische Communicatie (www.ogc.nl)
Cover design by Roos Meijboom

Evaluating and Mitigating Incentives for Risk Selection in Health Insurance Markets with Risk Equalization

Het evalueren en verminderen van selectieprikkels in zorgverzekeringsmarkten met risicoverevening

Thesis

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

Prof. dr. ir. A.J. Schuit

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

Friday 25 September 2026 at 13.00 hrs
by

Michel Oskam
born in Spijkenisse, the Netherlands.

Erasmus University Rotterdam

DOCTORAL COMMITTEE:

Promotors:	Dr. R.C. van Kleef
	Prof. dr. F.T. Schut
Assessment committee:	Dr. F. Eijkenaar
	Prof. dr. J. de Jong
	Prof. dr. J. Boone

CONTENTS

CONTENTS

List of publications and submissions 10

Chapter 1: Introduction 13

1.1 Background 14

1.2 Evaluating incentives for risk selection 15

1.3 Strategies to mitigate incentives for risk selection 17

1.4 The central aim 19

1.5 Structure of this dissertation 25

PART I: EVALUATING RISK EQUALIZATION THROUGH INNOVATIVE METHODS

Chapter 2: Selection incentives towards individuals diagnosed with a new chronic disease 29

2.1 Introduction 31

2.2 New Contribution 33

2.3 Conceptual Framework 34

2.4 Context, Data, and Methods 37

2.5 Results 41

2.6 Discussion 47

2.7 Appendix 50

Chapter 3: Selection incentives that result from heteroscedasticity of residual spending 55

3.1 Introduction 57

3.2 Conceptual framework 58

3.3 Data and methods 62

3.4 Results 67

3.5 Size of the problem and potential solutions 77

3.6 Discussion 81

3.7 Appendix 83

PART II: SIMULATING STRATEGIES TO MITIGATE SELECTION INCENTIVES

Chapter 4: Improving diagnosis-based cost groups to reduce incentives for risk selection 87

4.1 Introduction 89

4.2 The Dutch DCGs: history and current practice 91

4.3 Data 92

4.4 Methods 93

4.5 Results 97
4.6 Discussion 138
4.7 Appendix 106

Chapter 5: Supplementing risk adjustment with high-risk pooling to reduce incentives for risk selection 113
5.1 Introduction 115
5.2 Theoretical Framework 117
5.3 Data and methods 120
5.4 Results 126
5.5 Discussion 134
5.6 Appendix 138

Chapter 6: Mitigating incentives for risk selection by relaxing premium-rate restrictions in the presence of risk equalization 149
6.1 Introduction 151
6.2 Cross-subsidization achieved by risk equalization 153
6.3 Setting, data, and methods 157
6.4 Results 162
6.5 Discussion and conclusion 169
6.6 Appendix 174

Chapter 7: Conclusion and discussion 177
7.1 Conclusions 178
7.2 Overarching strengths and limitations of this research 182
7.3 Discussing the implications of this research 183

References 191

General appendix: Translation of ICPC list 205

Summary 211

Samenvatting 221

Dankwoord 233

PhD portfolio Michel Oskam 239

LIST OF PUBLICATIONS AND SUBMISSIONS

Chapters 2 through 6 are based on the following articles:

CHAPTER 2

Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2026). Does Health-Based Prospective Risk Adjustment Adequately Compensate for Individuals Diagnosed With a New Chronic Disease?. *Medical Care Research and Review*, *83*(2), 128-143. https://doi.org/10.1177/10775587251378167

CHAPTER 3

Oskam, M., Van Kleef, R. C., & Douven, R. (2024). Heteroscedasticity of Residual Spending After Risk Equalization: a Potential Source of Selection Incentives in Health Insurance Markets with premium regulation. *The European Journal of Health Economics*, *25*(3), 379-396. https://doi.org/10.1007/s10198-023-01592-9

CHAPTER 4

Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2023). Improving Diagnosis-Based Cost Groups in the Dutch Risk Equalization Model: The Effects of a New Clustering Method and Allowing for Multimorbidity. *International Journal of Health Economics and Management*, *23*(2), 303-324. https://doi.org/10.1007/s10754-023-09345-0

CHAPTER 5

Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2025). Supplementing Risk Adjustment With High-Risk Pooling Using Historical Data for Identifying the High Risks. *Journal of Risk and Insurance*, *92*(1), 166-202. https://doi.org/10.1111/jori.12500

CHAPTER 6

Oskam, M., Van Kleef, R. C., & Van de Ven, W. P. (2026). To What Extent Does State-of-the-Art Risk Equalization Create Cross-Subsidies on a Competitive Health Insurance Market?. *Submitted for publication.*

1

INTRODUCTION

1.1 BACKGROUND

Health insurance serves as an important mechanism to enable financial access for individuals to, often costly, medical services. However, without regulatory intervention, the premium of a health insurance contract will be adjusted to the expected cost of that contract, which could lead to unaffordable premiums for high-risk individuals with substantial and/or costly care needs (Rothschild & Stiglitz, 1976; Diamond, 1992; Van de Ven & Ellis, 2000a). Therefore, regulators of health insurance markets with competing insurers typically rely on premium regulations (e.g., through community-rating, where all premiums for the same coverage with the same insurer must be equal, irrespective of the individual-level risk of health expenditure) to protect the affordability of coverage for high-risk individuals. In addition to such constraints to premium-setting, other typical regulations in health insurance systems with regulated competition (e.g., the Netherlands, Germany, Switzerland) include an insurance coverage mandate, an open enrollment policy, and a standardized basic benefit package (Enthoven, 1978; Enthoven & Kronick, 1989). The coverage mandate guarantees insurance uptake by consumers, even when – because of the premium regulations – the price of coverage for low-risk individuals is above their expected expenditure on health services. The open enrollment policy prevents insurers from rejecting unfavorable applicants, for example, those with a risk of health expenditure that significantly exceeds the regulated premium. Last, the standardization of coverage is implemented to ensure that a comprehensive set of benefits is included within coverage packages. The combination of these regulations is meant to facilitate access to sufficient insurance coverage for the entire population.

However, with the broad set of regulations, the inherent variability in expected health care expenditures among the different risk profiles within the population persists. The prohibition of premium differentiation and the obligation to accept all applicants for insurance create financial incentives for insurers to prioritize attracting low-risk, financially appealing individuals and to dissuade the enrollment of high-risk individuals. This fosters risk selection: actions by consumers and health plans to exploit unpriced heterogeneity in risk and disrupt natural pooling arrangements (Newhouse, 1996). Risk selection can adversely impact the functioning of the health insurance system. For instance, insurers might influence their plan design to steer the enrollment of consumers based on anticipated medical service usage. By refraining from contracting high-quality care providers typically used by high-cost groups like chronically ill patients, insurers could undermine care quality (Van Kleef et al., 2013a; Van de Ven et al., 2015). Risk selection may also induce price distortions and inefficient sorting of individuals across health plans: when high-risk individuals gravitate towards specific plans, premium variations might reflect selection effects rather than the quality of the insurance plan, distorting the sorting of consumers across insurance plans and distorting the level playing field among insurers (Glazer & McGuire, 2000; Van de Ven & Ellis, 2000a). Moreover, efficiency of production is hindered

if risk selection proves to be a more effective strategy for cost-control than aligning with consumer preferences (Van de Ven & Ellis, 2000a; Glazer & McGuire, 2000). These dynamics resulting from actions of risk selection could seriously undermine the functioning of the health (insurance) system, threatening quality, efficiency, and accessibility (Einav & Finkelstein, 2011; Van de Ven et al., 2015).

To mitigate incentives for risk selection in regulated health insurance markets, regulators have implemented an additional instrument: risk equalization. Through this instrument, also known as risk adjustment (the terms are used interchangeably throughout this dissertation), competing insurers are compensated for the predictable variation in health spending of their enrollees (Ellis et al., 2018). Through a predefined set of risk factors (risk adjusters), a prediction is made of health spending per individual. Over the past decades, risk equalization models have evolved from simple demographic models – based on the age and sex of individuals – to sophisticated health-based models, including risk adjusters based on hospital treatments and pharmaceutical usage in prior years and past health spending. Through these comprehensive models, better estimations of individual health spending can be made to compensate for predictable variation in health spending (Van Kleef et al., 2019; McGuire et al., 2021a). Therefore, the predictable loss to insurers for contracting high-risk consumers (for instance, elderly individuals with a chronic heart condition) is reduced. However, although reduced, the leftover predictable loss still provides insurers with incentives to avoid the enrollment of chronically ill individuals, despite the presence of a sophisticated risk equalization model (Van Kleef & Van Vliet, 2022b). Recent studies have indicated that in countries with sophisticated risk equalization (e.g., the Netherlands, Germany, and the United States), health insurers still engage in actions of risk selection to deter unprofitable consumers from enrollment (Bauhoff et al., 2017; Decarolis & Guglielmo, 2017; Shepard, 2022; Stolper et al., 2022). The undercompensating (overcompensating) of various (specific) risk groups in the population therefore remains an important focus for improvement in risk equalization. To protect access to high-quality care for vulnerable, high-risk individuals and the functioning of the health system altogether, further reduction of risk selection incentives is crucial.

1.2 EVALUATING INCENTIVES FOR RISK SELECTION

Many studies have been conducted on the effectiveness of risk equalization by quantifying the risk selection incentives that remain after risk equalization payments. Typically, such studies find unhealthy or, alternatively, chronically ill individuals to be structurally undercompensated by the model. To provide a valuable evaluation of the functioning of the risk equalization model, the predictable profit or loss for relevant subgroups is quantified as an approximation of selection incentives for insurers towards the identified subgroups. To identify these subgroups within the population, external data (i.e., data not

included in the risk equalization model that is evaluated) are required. This is because the property of the estimation method used for calibrating the model (ordinary least squares regression) ensures that the variation in health spending *between* the risk adjusters used in the model is completely compensated. In other words, any defined subgroup based on any specific risk adjuster from the model will – on average – be compensated completely for their expected health expenditure, meaning there is no predictable profit nor loss. Therefore, to facilitate an effective evaluation of risk selection incentives after risk equalization, subgroups of the population should be derived from external data to gain insight into their respective results after risk equalization payments (Van Veen et al., 2015).

Moreover, the true health status of individuals extends beyond the risk adjusters included in the model. Therefore, external indicators are expected to capture different subgroups than (combinations of) the risk adjusters included in the model. For instance, information derived from health surveys has been used for the evaluation of risk equalization by Withagen-Koster et al. (2018) and Withagen-Koster et al. (2020). The authors find that, for the identified subgroups of (un)healthy individuals, Dutch risk equalization leaves nontrivial predictable profits and losses. However, the self-reported information in these surveys inherently suffers from subjectivity, in addition to the fact that the size is relatively limited (N = 380k). Considering these limitations, the availability of a comprehensive set of data on patient records from Dutch general practitioners (GPs) presents a new opportunity to evaluate the functioning of the Dutch risk equalization model. The sophisticated data concern the diagnostic information from the Nivel Primary Care Database (Nivel-PCD), which indicates the presence or absence of 683 diagnoses for all individuals included in the data (N = 1.5m) during the years 2015 through 2019 (Nivel, 2019; Vanhommerig et al., 2025). Out of the 683 diagnoses, 109 have been labeled as chronic diseases and thus provide the opportunity to identify specific subgroups of chronically ill individuals (Nielen et al., 2019). Within the context of the Dutch health system, data acquired from the patient records of GPs is of significant relevance: GPs serve as gatekeepers within the health system, providing valuable insight into the health status of individuals, as most will have passed the GP before making use of other health services (e.g., hospitals or clinics). Additionally, providers of other health services (e.g., medical specialists) also refer patients back to their GPs. Therefore, the objective health status of chronically ill individuals is typically diagnosed and registered by GPs. Moreover, nearly all inhabitants of the Netherlands are registered with a GP. Through this dataset, individuals can be selectively grouped based on the presence of chronic diseases to evaluate the functioning of the RE model for these vulnerable risk types. So, within this dissertation, the Nivel-PCD data is used to quantify the functioning of the Dutch risk equalization model.

The aim of this dissertation is to contribute to the existing body of academic literature on risk equalization and how its functioning is evaluated. The common method to evaluate the extent to which (alterations to) risk equalization reduce(s) incentives for risk selection

is to simulate the predictable profits and losses for selective subgroups of consumers, derived from external data, for a single year (for instance, Van Kleef et al., 2019; Zink & Rose, 2021). However, the effectiveness of this method is notably subject to the data used to identify subgroups of consumers. For instance, if the aim is to identify all individuals with a specific chronic disease, information from a health survey or registered pharmaceutical usage may not be able to identify the complete group of individuals diagnosed with the disease in question. Not all who suffer from the same disease follow the same care trajectory (e.g., in the (dosage of) pharmaceuticals used), leading to difficulties in identifying all individuals that should qualify. To effectively evaluate risk equalization through quantifying the selection incentives towards a subgroup of the population, the adequate identification of individuals is crucial. As argued above, through the Nivel-PCD, a stronger and more objective identification of chronically ill individuals can be achieved (Van Kleef & Van Vliet, 2025).

Moreover, it could be argued that the focus on a single year in the evaluation of risk equalization is shortsighted. Although consumers are free to annually switch their insurer or insurance plan, selection incentives experienced by insurers will likely be influenced by more than the expected result for a single year. For instance, annually recurring predictable financial losses may compile, amplifying the selection incentives. Therefore, a multi-year approach in the evaluation methodology should be considered. Last, the typical quantifying of selection incentives through predictable profits and losses is only based on the average result for the identified subgroup. However, this method does not consider important underlying information. The variation in the predictable profits and losses within the defined group is valuable information, as it may contribute to the selection incentives experienced by insurers. Two subgroups with an equal average predictable loss could still present different selection incentives to insurers through differences in the variation around that average result. For example, a greater variation surrounding the same average may be unappealing to risk-averse insurers. Therefore, this component should be incorporated in the evaluation of risk equalization, in addition to the multi-year approach. As such, this dissertation builds on the academic literature on risk equalization by using the GP data to critically and innovatively evaluate the functioning of Dutch risk equalization.

1.3 STRATEGIES TO MITIGATE INCENTIVES FOR RISK SELECTION

Provided that the problematic undercompensation of chronically ill individuals remains present in risk equalization, improvement to the system remains necessary. In competitive, regulated health insurance markets that include risk equalization, three typical, complementary strategies exist to reduce the risk selection incentives that remain. The first concerns the modification of the risk equalization model: either by adding new risk

adjusters, expanding the existing ones, altering the estimation method, or any combination of these (e.g., Eijkenaar et al., 2018; Schindler et al, 2025). The gradual improvement of risk equalization models over the years has followed this approach, including the addition of new risk adjusters when new information of predictive value became available. However, now that risk equalization models are increasingly comprehensive, the potential for (new) risk adjusters from new sources of predictive information is limited. Leftover predictive data sources that cover the entire population are scarce and may not be suitable for uptake into the model. As an alternative, recent studies have explored the potential for alternative estimation methods to derive the payment weights of the risk equalization model (e.g., constrained regression) (Van Kleef et al., 2017; Renker et al., 2025). A concern with this approach is that, although better outcomes may result, the complexity of the model increases, and interpretability of the outcomes diminishes.

The second strategy to reduce incentives for risk selection is to implement risk-sharing as a supplement to risk equalization. Through risk-sharing, insurers are compensated for (a portion of) the health spending of their enrollees (after risk equalization payments), reducing their financial risk and the predictable profits and losses after risk equalization (McGuire & Van Kleef, 2018b). However, while incentives for risk selection are reduced through risk-sharing, incentives for cost-control for insurers are reduced as well. For instance, by applying proportional risk-sharing to compensate for 50% of health spending after risk equalization, predictable profits and losses (and thus selection incentives) are reduced by 50%, but so are incentives for cost-control. With insurers less responsible for efficient financial behavior, health spending within the system could inflate. When insurers receive more financial compensation when the spending of their enrollees increases, the need to arrange efficient contracts for medical services diminishes. Therefore, this strategy presents an important tradeoff between selection incentives and efficiency or cost-control.

Finally, the third strategy that could be considered to reduce the incentives for risk selection is to relax premium-rate restrictions. By allowing insurers to differentiate their premiums to reflect the predictable leftover variation in health spending after risk equalization payments, incentives for risk selection are mitigated. However, in that scenario, specific high-risk individuals may be charged unaffordable high premiums, threatening affordability and diminishing solidarity. Therefore, all three strategies present a tradeoff to regulators attempting to reduce incentives for risk selection. The implications of these strategies are typically addressed in research when applied in isolation, while the strategies could also be implemented complementarily. Careful balancing of the strategies could potentially lead to a desirable outcome for policymakers regarding the various tradeoffs concerning risk selection, practicality of the model, efficiency, and affordability.

Within this dissertation, variations of the discussed strategies are studied to quantify their respective effectiveness at reducing incentives for risk selection. In addition to the use

of the described GP data to quantify the incentives for risk selection that exist after risk equalization, the availability of the data provides an opportunity to measure the effectiveness of the strategies simulated within this dissertation to reduce them. Thus, within this dissertation, the GP data by Nivel-PCD is used to evaluate potential adaptations or supplements to risk equalization and to develop and test new methods for evaluating the effectiveness of risk equalization.

1.4 THE CENTRAL AIM

The aim of this dissertation is to develop new evaluation measures for selection incentives that remain after risk equalization in social health insurance markets and test strategies to further reduce those incentives. The dissertation is structured in two main parts. The first examines two innovative methods for evaluating the functioning of risk equalization systems, seeking to measure risk selection incentives through both GP data and administrative data. The second part of the dissertation examines three strategies to improve the functioning of risk equalization or reduce risk selection incentives through further supplements to the insurance system. One concerns the adjustment of an existing risk adjuster within the model, while the second applies risk-sharing. The effect of these two strategies is evaluated with the GP data made available for this dissertation. The third strategy covers a critical assessment of the need for premium-rate restrictions in Dutch health insurance by simulating a scenario of risk-rated premiums after risk equalization. In doing so, the three strategies to mitigate risk selection incentives as discussed in section 1.3 are addressed and simulated within this dissertation.

In sections 1.4.1 and 1.4.2, the two parts of the dissertation are described in greater detail, as well as how they both contribute to the existing body of research on risk equalization.

1.4.1 Evaluating risk equalization through innovative methods

This part of the dissertation is focused on deriving and applying innovative methods to quantify risk selection incentives after risk equalization to gain a better understanding of the problem. Dividing the population into subgroups vulnerable to risk selection is important, but the subgroups are likely to suffer from internal heterogeneity: for instance, not all patients with heart diseases are equal. One crucial dimension of heterogeneity for such a relatively static classification of individuals is the time passed since a diagnosis was first made. This subdivision can be particularly relevant for risk equalization systems of prospective nature, where health spending is predicted using health care utilization in preceding years. In the scenario where an individual develops a chronic disease, the prospective risk equalization is likely unable to compensate adequately for the unpredictable rise in health spending in that year. As a result, while the overall compensation for a group of patients with heart diseases may be considered adequate, the compensation

for the specific subgroup of newly diagnosed individuals is expected to be inadequate. Therefore, the first research question of this dissertation is:

Question 1: Does health-based prospective risk equalization adequately compensate insurers for individuals diagnosed with a new chronic disease?

Inadequate compensation for individuals who develop a new chronic disease can be problematic in terms of risk selection. Its magnitude depends on whether insurers can steer the behavior of these individuals in their choice of insurance product. Moreover, the potential for this distortion of enrollment in turn depends on whether, when, and how the individual who develops a new disease responds to their change in health status. Various studies have documented the relation between health plan switching and its relation to health status, indicating that individuals who incur a health shock are likely to reconsider their choice of insurance plan (Cutler & Zeckhauser, 2000; Bolhaar et al., 2012; Einav et al., 2013; Croes et al., 2025). Therefore, the contribution of this research is to study the dynamic between a change in health status, variation in health spending and compensation, and insurance enrollment behavior within the context of risk selection. Through the combination of multiple years of the GP data, the first step of this process is to identify individuals that are diagnosed with a new chronic disease. After, the risk equalization models for the year of diagnosis and the two years before and after diagnosis are simulated to track the evolution in health spending, risk equalization payments, and the resulting residual spending. Finally, the behavior of these individuals in terms of insurer switching is studied to indicate whether and when individuals respond to the change in health status. If the newly diagnosed individuals are undercompensated in the years surrounding diagnosis and take this health change into account when switching their insurance plan, insurers may try to deter their enrollment. As discussed before, the potential resulting actions of risk selection could threaten the functioning of the health system.

While the study above strays from traditional research on the evaluation of risk equalization functioning – by identifying a specific part of subgroups and tracking the residual spending over five years –, the underlying assumption of evaluation studies remains the same. It is typically assumed (either explicitly or implicitly) that in the absence of predictable profits and losses, there is no financial incentive for insurers to distort the natural enrollment of individuals (Glazer & McGuire, 2000; Van de Ven & Ellis, 2000a). However, even when risk equalization perfectly reduces the average predictable profits and losses across all different risk types to zero, selection incentives may persist. This is attributed to varying distributions of residual spending across the risk types, where – despite the equal average residual spending – groups with higher predictable health spending typically have greater variance in residual spending. As a result of this heteroscedasticity (the difference in the variance in residual spending between subgroups of consumers), insurers typically face more uncertainty about their financial results after risk equalization

when contracting individuals with relatively high health spending. If an insurer would contract an individual from a subgroup with relatively high variance in residual spending and another from a subgroup with less variance, the average expected financial result on both contracts after risk equalization may be zero, but the certainty of the two outcomes differs. Therefore, the variance in residual spending between subgroups of consumers might lead to selection incentives if the uncertainty is undesirable to insurers. This study seeks to quantify the variance in residual spending between subgroups of consumers as a new source of risk selection incentives through the following second research question:

Question 2: What does the heteroscedasticity of residual spending look like, and how can it affect selection incentives?

Within the study, two hypotheses are studied about how heteroscedasticity can lead to new incentives for risk selection. First, the uncertainty in residual spending is linked to the observation that insurers are typically risk averse: if the outcome to insurers is more uncertain for an enrollee from a specific subgroup with high variance in residual spending, insurers will likely require an additional payment to cover the risk of high potential losses. Second, insurers are subject to regulations regarding solvency that require them to maintain larger amounts of capital to cover larger degrees of uncertainty on their financial outcomes. In a market free of premium-rate restrictions, the uncertainty in residual spending and the need for elevated capital preservation will likely be reflected in the premium through mark-ups. In regulated markets with premium regulations, such differentiation in mark-ups is impossible, potentially leading to selection incentives. The simulated mark-ups therefore serve as an indication of selection incentives in regulated markets that include premium regulations. Therefore, the proposed innovative strategy for risk equalization evaluation identifies a new source of risk selection incentives and illustrates how it could be counteracted. The combination of the two studies of this part of the dissertation sheds new light on risk selection incentives in social health insurance markets that remain after risk equalization.

1.4.2 Simulating strategies to mitigate selection incentives

This next part of this dissertation investigates strategies to mitigate selection incentives that remain a threat to the functioning of the health (insurance) system. Further advancement to the risk equalization framework is required, as discussed in section 1.3. Therefore, the third study of this dissertation examines a specific modification of the risk equalization model by evaluating a recent modification to the risk adjuster 'Diagnosis-based Cost Groups' (DCGs) of the Dutch risk equalization model. The risk adjuster is designed to identify individuals with chronic conditions based on hospital treatments. Recognizing that conditions extend beyond treatment alone and not all conditions are captured within a DCG, this study evaluates the effect of the modification on the functioning of the risk equalization model. To do so, the study leverages the external GP data to identify

individuals diagnosed with a chronic disease by their GP. Through this method, the aim is to evaluate how well the updated risk equalization model recognizes the presence of chronic illness in its compensations. The third research question of this study seeks to investigate this nuanced aspect:

Question 3: To what extent do modifications to the Dutch DCG-classification reduce selection incentives?

In addition to the common method for enhancing risk equalization through modification of the included risk adjusters, complementary strategies are often used. One strategy is to implement risk-sharing, which can be designed in various modalities to reduce financial risk for insurers. The challenge within risk-sharing design, however, is how to achieve the largest reduction in selection incentives while compromising the fewest incentives for cost-containment. Of the possible modalities of risk-sharing, high-risk pooling represents a specific example to effectively reduce selection incentives. Through high-risk pooling, insurers are compensated for (a share of) the health spending (after risk equalization payments) for a predefined set of high-risk individuals (Van Barneveld et al., 1998). Therefore, high-risk-sharing is conditionally applied to a subset of individuals that are expected to be predictably unprofitable to insurers, rather than compensating for spending of all individuals above any threshold, for instance. In the latter scenario, insurers lose the incentives for cost-containment when the threshold approaches. Importantly, despite the idea of high-risk pooling being almost 30 years old and its described benefits over other risk-sharing modalities, it is not applied in health insurance systems. A vital part of the design of the high-risk pooling modality, however, is how to effectively identify the individuals that are expected to be (substantially) unprofitable to insurers.

Within this context, the availability of extensive health data and advanced analytical techniques offers a promising avenue for identifying individuals of high risk more effectively to set up a high-risk pooling modality (McGuire & Van Kleef, 2018b). This study specifically investigates the potential of high-risk pooling as a strategy to mitigate remaining selection incentives within health insurance systems with risk equalization. By leveraging the available datasets on Dutch risk equalization models from multiple years and machine learning algorithms, this study aims to refine the identification process for high-risk individuals to simulate a high-risk pooling modality. The fourth research question of this study is therefore designed to explore how high-risk pooling can effectively complement existing risk equalization and address selection incentives that prevail:

Question 4: To what extent does high-risk pooling reduce the selection incentives that remain after sophisticated risk equalization?

The increased access to extensive data concerning health and health care utilization has facilitated the improved identification of high-risk individuals, benefiting both regulators and insurers. From the perspective of risk selection, it is crucial that this group of insured individuals is identified and their spending is appropriately compensated. The information used to estimate risk equalization models holds significant explanatory power to identify such risk, particularly through information on health utilization and health spending. Moreover, due to unsuitability for direct uptake, most of the information used for estimating the model for one year is not used for subsequent models, despite its predictive value. Therefore, it is prudent to leverage this available data in alternative ways to identify high-risk individuals and further diminish incentives for risk selection within the health insurance market through high-risk pooling.

The final strategy to mitigate risk selection incentives as discussed in section 1.3 concerns the relaxation of premium-rate restrictions. This strategy reduces regulation-induced risk selection through the modification of regulation to reduce these incentives rather than counteracting the incentives for risk selection stemming from the prevailing regulation (section 1.1). Complete relaxation of the premium-rate restrictions without the presence of risk equalization removes risk selection incentives altogether but can lead to problems of unaffordability of coverage. The motivation for regulators behind enforcing premium-rate restrictions is to safeguard access to coverage through realizing implicit cross subsidies from low-risk individuals to high-risk individuals. However, provided that the quality of risk equalization has significantly advanced since implementation, a notable share of cross-subsidization is generated explicitly through the instrument. Therefore, it is interesting to evaluate to what extent sophisticated risk equalization is able to realize the cross subsidies that are desired through premium-rate restrictions. This leads to the final research question of this thesis:

Question 5: To what extent does state-of-the-art risk equalization create cross subsidies in a competitive health insurance market?

It is well understood that risk equalization, despite the level of sophistication, does not compensate for all variation in health spending. Therefore, releasing premium-rate restrictions, under the assumption that risk equalization creates cross subsidies, directly leads to premium differentiation based on the predictable profits and losses that remain after compensation. While this threatens accessibility to coverage, incentives for risk selection are avoided, proposing an interesting tradeoff. Or alternatively, this analysis raises the question of whether, in the presence of state-of-the-art risk equalization, the marginal benefits of premium-rate restrictions (the additional cross-subsidization on top of the cross-subsidization achieved by risk equalization) outweigh the marginal costs (the potential effects of regulation-induced selection). The third strategy differs from the

other two in that it is focused on adapting the regulation itself rather than counteracting the incentives stemming from it.

1.4.3 A note on the data sources used for empirical analyses

To answer the research questions outlined above and facilitate an adequate evaluation of the Dutch risk equalization model, the discussed GP data is combined with another comprehensive dataset that has been made available for this study: the administrative data used to estimate the Dutch models that were in place for 2018 through 2022[1]. The data cover information on health spending and the risk adjusters included in the model for all individuals enrolled in the mandatory Dutch basic health insurance system during the selected period (N = 17m). The model is considered state-of-the-art with a broad range of risk adjusters, including demographic, socioeconomic, and health-based variables. The model is annually updated, with risk adjusters expanded, added, or replaced, and changes implemented to risk-sharing modalities and the estimation method. For the actual design of the Dutch risk equalization model per year since 2011, see Zorginstituut Nederland (2026).

Replicating the model offers the opportunity for an external evaluation through combination with the GP data from Nivel-PCD. The GP data from Nivel-PCD, as explained in section 1.2, contains the diagnostic information on 683 conditions for over 1.3 million individuals from 2015 through 2019. This information is coded according to the International Classification of Primary Care (ICPC), *version 1*, by the Wonca International Classification Committee. For an overview of the complete list, see Het Nederlands Huisartsen Genootschap (2024), and for more information on the data collection by Nivel, see Nivel (2025).

It should be noted that the GP data is available for roughly ten percent of the entire population. Through a data-manipulation technique (the raking method (Battaglia et al., 2009)), a rebalancing weight is generated to ensure the prevalences in the subset match those of the total population. While the technique is insufficient to implement the incomplete GP data directly into the risk equalization model – and ensure an (on average) accurate compensation for the defined groups –, the identification of selective subgroups for purposes of evaluation is extremely beneficial to the functioning of the system, nonetheless.

1 The estimation of the risk equalization model for year *t* (e.g., 2022) is performed using data on health spending from year *t-3* (e.g., 2019). Therefore, the underlying data of the models for 2018 through 2022 correspond with the GP data of the respective years 2015-2019.

1.5 STRUCTURE OF THIS DISSERTATION

The chapters of this dissertation follow the structure outlined in section 1.4. The first part examines the innovative evaluation strategies for risk equalization as discussed in section 1.4.1. Chapter two answers the first research question to quantify the functioning of the Dutch risk equalization model for individuals who develop a chronic disease. Chapter three answers the second research question by quantifying how heteroscedasticity in residual spending after risk equalization leads to incentives for risk selection. Part two of this dissertation serves to simulate three strategies to mitigate risk selection incentives in the Dutch risk equalization model, as discussed in section 1.4.2. Chapter four answers the third research question by examining how an improvement to one of the risk adjusters in the risk equalization model leads to better compensation of chronically ill individuals. In chapter five an answer is provided to research question four of this dissertation by simulating and evaluating an innovative modality of risk-sharing to reduce incentives for risk selection. In chapter six, research question number five is answered. This chapter seeks to challenge the regulation regarding premiums in the presence of sophisticated risk equalization to mitigate selection incentives in health insurance.

The two parts are designed to address knowledge gaps in literature on the respective subjects and contribute to the overall body of academic research. Last, the findings of chapters two to six are summarized and discussed in greater detail. The implications of the findings of this dissertation are then discussed for both policymakers and for further research.

PART I

Evaluating Risk Equalization
through Innovative Methods

2

SELECTION INCENTIVES TOWARDS INDIVIDUALS DIAGNOSED WITH A NEW CHRONIC DISEASE

Based on: Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2026). Does Health-Based Prospective Risk Adjustment Adequately Compensate for Individuals Diagnosed with a New Chronic Disease? *Medical Care Research and Review, 83*(2), 128-143. https://doi.org/10.1177/10775587251378167

Abstract

Many regulated health insurance markets use prospective risk adjustment (RA) to mitigate risk selection incentives for insurers. However, prospective RA might underpay insurers for people diagnosed with a new chronic disease. By tracking spending and RA payments over the period *t–2* to *t+2* for individuals diagnosed with a new chronic disease in year *t*, we find a substantial payment gap in year *t* and, to a lesser extent, in prior and/or subsequent years. The extent to which these gaps stimulate selection incentives for insurers depends on the possibilities for insurers to distort consumers' choice of insurance products. Possibilities which – in turn – depend on whether and when consumers respond to the onset of the chronic disease when choosing an insurance product. By analyzing 'insurer switching' in the period *t–2* to *t+2*, we find that – on average – people first diagnosed with a chronic disease are more likely to switch insurers than others.

2.1 INTRODUCTION

Around the world many health insurance markets are based on principles of regulated competition (McGuire & Van Kleef, 2018a). Examples include the basic health insurance schemes in the Netherlands, Germany, and Switzerland; the Marketplaces in the United States under the Affordable Care Act, as well as Medicare Advantage. Economic theory suggests that a balanced combination of competition and regulation can enhance the efficiency and fairness of health insurance markets (Enthoven, 1978; Enthoven, 1988b). Typical aspects of regulation include standardization of coverage, open enrollment requirements, and premium rate restrictions. However, while such regulations enhance the accessibility and affordability of health insurance for individuals with a high risk of medical spending, they can exacerbate selection problems (Van de Ven & Ellis, 2000a). For example, mandating insurers to charge identical premiums to individuals for an insurance contract, irrespective of age, health status, and prior health spending, creates predictable profits and losses, which confront insurers with incentives for risk selection (Glazer & McGuire, 2000; Newhouse, 1996; Pauly & Langwell, 1983; Van de Ven et al., 2017). Possible selection actions range from subtle marketing strategies to distortions of benefits, cost-sharing schedules and provider networks. For example, health insurers might choose not to contract high-quality providers that treat disproportionate shares of unprofitable risk types (Bauhoff et al., 2017; Decarolis & Guglielmo, 2017; Shepard, 2022; Stolper et al., 2022; Van de Ven et al., 2015). These actions could seriously undermine the functioning of health (insurance) systems through skimping on quality and access to health care (Einav & Finkelstein, 2011; Van Kleef et al., 2024a).

To mitigate risk selection by insurers, regulators have implemented schemes of prospective risk adjustment (RA) to compensate insurers, based on estimations of predictable variation in health spending between individuals (Ellis et al., 2018). Over the past decades, prospective RA models have evolved from simple demographic models to sophisticated health-based models with risk adjusters based on diagnostic information, socioeconomic variables, and/or prior health care utilization/spending. The uptake of health information has significantly improved the predictive strength of RA models, leading to more accurate compensation of insurers and reducing incentives for risk selection (Layton et al., 2018c; Van Kleef et al., 2018a).

Despite the improvements, it is questionable whether prospective RA models adequately compensate insurers for a specific group of enrollees: those who develop a (chronic) disease. The reason is threefold. First, individuals who develop a disease are not immediately diagnosed. Initial examinations and treatments may precede the eventual diagnosis and lead to elevations in health care utilization and spending. Therefore, for as long as the diagnosis has not been registered, the RA model might not adequately or timely acknowledge the rise in health spending through higher compensations. Second, in prospective

RA health spending of individuals is predicted using health care utilization in the preceding year, creating a time lag in the compensation of insurers (e.g., diagnoses from year *t–1* are used to predict health spending in year *t*). As a result, a diagnosis of a chronic disease made in year *t* – and its related inflated health spending – will be recognized by the RA model in year *t+1* at the earliest, leaving a potential financial loss to insurers in the years prior. Third, in the first years after diagnosis, individuals who develop a chronic disease in general require more examinations and treatments than those for whom the disease has stabilized over time. So, within the group of individuals with a chronic disease, those recently diagnosed might be, on average, costlier to insurers than the rest of the group and yield inadequate compensation.

In this paper, we track the performance of prospective RA for the group of individuals diagnosed with a (new) chronic disease in year *t*. More specifically, the first goal of this study is to calculate the mean financial result after prospective RA in years *t–2, t–1, t, t+1,* and *t+2* for individuals who are diagnosed with a new chronic disease in year *t*.

The extent to which a negative or positive financial result for insurers regarding consumers who are diagnosed with a new chronic disease is problematic depends on whether insurers are able to steer the behavior of those consumers in the choice of insurance product. In turn, the potential to distort the decision-making depends on whether, when, and how the individual in question responds to the change in health status. So, the more consumers let the change in health status reflect in their choice of insurance plan *before* a year in which they are unprofitable to insurers, the stronger will be the incentive for insurers to manipulate their enrollment. Therefore, to expand on the first goal of this study, we track the consumer response of the group of individuals diagnosed with a new chronic disease in year *t*. More specifically, the second goal is to track insurer-switching behavior[1] in year *t–2* to year *t+2* of individuals diagnosed with a new chronic disease in year *t*.

Our study is based on the Dutch basic health insurance. This health insurance is mandatory for every person who lives or works in the Netherlands (over 18 million people in 2025). The benefit package is standardized and includes primary care, inpatient and outpatient hospital care, and pharmaceutical care, among other services. Each year consumers can switch insurance plans. Although the benefit package is standardized, plans can differ in terms of provider network and coverage for out-of-network spending. Through selective contracting, preferred provider contracts, and supplementary insurance plans, competing insurers can differentiate themselves to attract enrollment. Plans with more generous networks are typically more expensive than restricted plans, but through the

1 We use insurer switching as a proxy for the consumer response. This period includes four annual switching windows. In the 'Methods' section we provide more details on the operationalization n of this proxy, and we elaborate on the consequences of its use in the Discussion.

open enrollment policy for every new calendar year, insurers are mandated to accept any applicant for insurance.

The starting point of our empirical analysis is the Dutch RA model, a sophisticated RA model that includes a broad spectrum of risk adjusters based on age, sex, health, and socioeconomic variables. The health indicators used in the Dutch model are of a prospective nature. Although the empirical findings are specifically relevant for the Netherlands, our conceptual points are potentially relevant for any country that relies on a prospective RA formula to mitigate selection problems in regulated health insurance markets.

This paper is organized as follows. The 'New contribution' section explains how this study provides new insights in relation to existing literature. In section 2.3, 'Conceptual framework', we explain how gaps between spending and RA payments for individuals diagnosed with a new chronic disease can be problematic in light of risk selection. Next, we describe the Dutch context and the data and methods used for our analysis. We then present the main findings. Finally, we discuss the implications of our findings for the design of health insurance markets.

2.2 NEW CONTRIBUTION

Over the past decades, many papers have been published on the design and evaluation of RA in health insurance markets. A typical method to evaluate the extent to which an RA model reduces selection incentives is to simulate the financial gaps after RA for selective groups of (un)healthy enrollees (see, for instance, McGuire et al., (2021a); Van Kleef et al., (2019); Zink and Rose (2021). Ideally, these evaluations focus on groups that are vulnerable to selection actions by insurers. For example, in their evaluations of American RA models, Montz et al. (2016) use mental health disorder diagnoses, and Geruso et al. (2019) use prescribed drugs to identify subgroups that are potential selection targets. Other researchers used diagnostic information from electronic patient records (e.g., Oskam et al., 2023; Van Kleef et al., 2020a) or health survey information to identify individuals with specific chronic diseases (e.g., McWilliams et al., 2012; Withagen-Koster et al., 2023).[2] Although the subdivisions of the population in these studies provide a meaningful basis for evaluating selection incentives, such subgroups defined by the presence of a chronic disease can still be relatively heterogeneous. One dimension of heterogeneity is the time passed since a diagnosis was first registered. This study evaluates the outcomes of a sophisticated RA model over a 5-year period for individuals who are diagnosed with a new chronic disease, setting it apart from prior studies.

2 For more examples on risk adjustment evaluation studies, see: Bergquist et al. (2019); McGuire et al. (2021b); Xu et al. (2021).

In addition to measuring the RA performance, this study examines the insurer-switching behavior of individuals who are diagnosed with a new chronic disease. Over the past decades, several papers have been published on health plan switching and its relation to health status (Boonen et al., 2016; Cunningham & Kohn, 2000; Ellis et al., 2017; Nuscheler & Knaus, 2005; van Vliet, 2006). Individuals who (anticipate to) incur a health shock are likely to reconsider their insurance plan (Croes et al., 2025; Cutler & Zeckhauser, 2000; Einav et al., 2013; Ellis, 1989; Robinson et al., 1993; Van de Ven & Van Praag, 1981). For instance, in a study on Chinese health insurance, Xu et al. (2016) have found increased uptake of insurance for individuals who had recently fallen ill or developed a chronic illness. Similarly, a positive association was found between lagged utilization of services by a general practitioner (GP) or a specialist and the purchase of supplementary private health insurance in Ireland (Bolhaar et al., 2012). Lange et al. (2017) have found evidence of a higher propensity to renew supplementary hospital insurance in Germany for individuals with a below-average health status, and Bajari et al. (2014) have found that individuals in more generous insurance plans face a greater risk of a health shock than others. It should be noted that these findings are dependent on the context and can vary from system to system.[3]

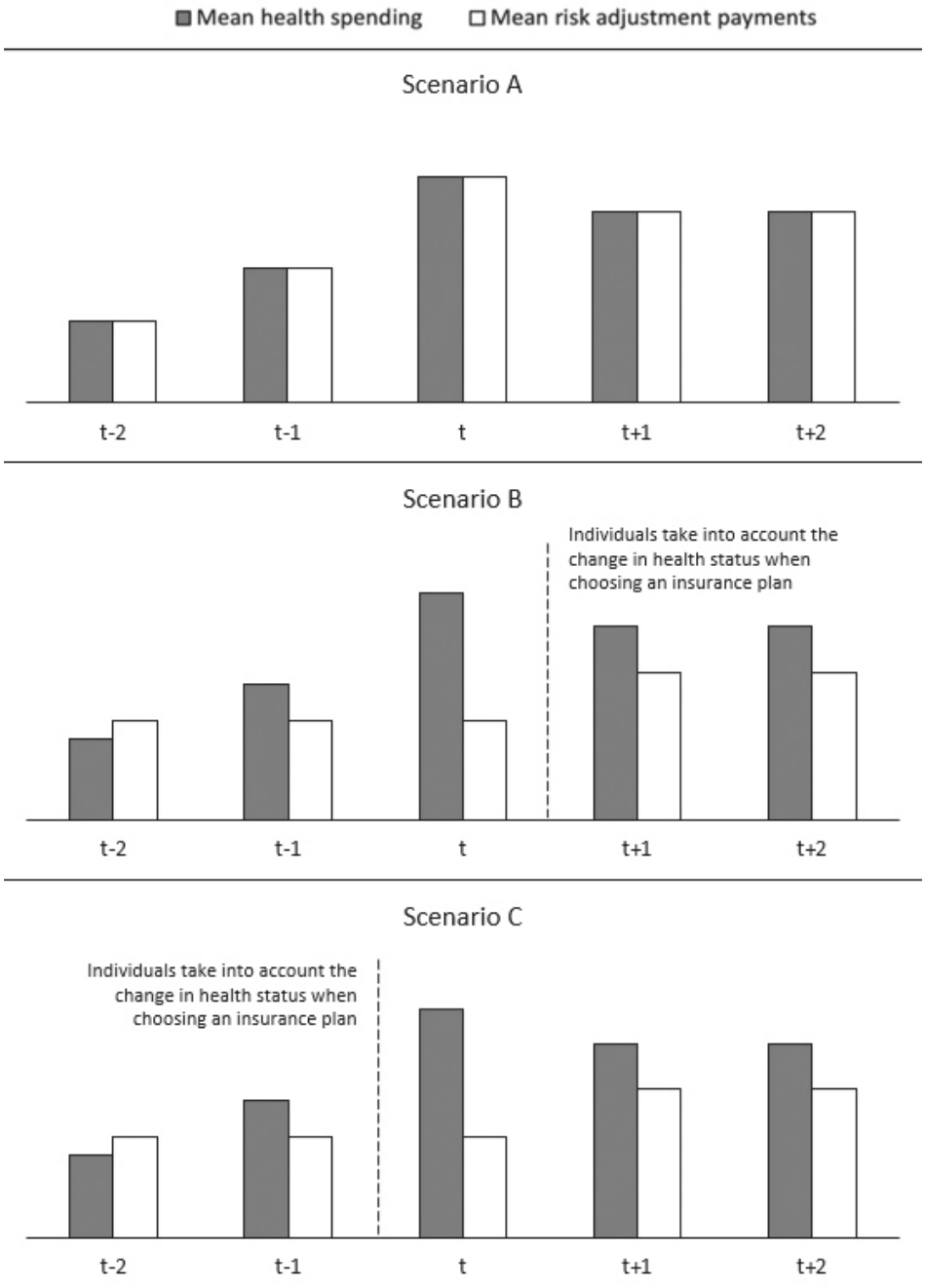
These differences in the latent health status of enrollees between plans provide an indication of consumer behavior in anticipation of, or in response to, a change in health status. Our study adds to this literature by examining the switching behavior of individuals who are diagnosed with a new chronic disease. We go beyond the existing literature by combining the consumer response in terms of insurer switching of this group with multi-year patterns in spending and RA payments. To the best of our knowledge, this focus has not been applied in earlier research. In the conceptual framework, we elucidate why the combined approach is crucial for evaluating the functioning of the health insurance market.

2.3 CONCEPTUAL FRAMEWORK

This section elaborates on the two main factors that determine incentives for insurers to select against individuals diagnosed with a new chronic disease: (1) the outcomes of the RA model and (2) the response of these individuals to their change in health status. This can be illustrated with Figure 2.1, which presents a simplified hypothetical scenario of the evolution of health spending over 5 consecutive years for a group of people who are newly diagnosed with a chronic disease in year *t*. In the three panels of the figure, the pattern in mean health spending is identical: before the diagnosis is first registered (year *t*) mean spending starts to rise (e.g., due to the first signals of the upcoming disease); in year *t* mean spending peaks (e.g., due to extensive health checks and treatments); after

3 For example, systems with multiple, largely different insurance coverage options (e.g., in the United States) will have notable inter-system switching in addition to plan switching within the system.

Figure 2.1. Three hypothetical scenarios of mean health spending, risk adjustment payments, and switching patterns over a five-year period for individuals who are diagnosed with a new chronic disease in year *t*.



Note: For simplicity, premiums are ignored in the figure. As a result, all financial compensation received by insurers is through RA payments.

year *t* mean spending stabilizes, though at a higher level than before the onset of the disease (e.g., due to structural treatment).

A vital first question on selection incentives for insurers is how RA payments develop over the considered 5-year period. For illustrative reasons, Figure 2.1 distinguishes three scenarios. In Scenario A, the mean RA payment for the group of interest perfectly follows the mean spending. In this scenario, there are no (predictable) profits and losses for this group, implying the absence of selection problems regarding this group.

However, it is unlikely that RA payments perfectly match spending for the group of interest in reality. Scenarios B and C demonstrate notable gaps between health spending by consumers and the compensation received by insurers. As a result of the belated initiation of compensation through prospective RA, a significant payment gap is expected in year *t*. For year *t–2*, when consumers are still healthy, we hypothesize an overpayment. In year *t–1* (when signals of the chronic disease start to occur) and years *t+1* and *t+2* (when the chronic disease is present, though stable), we hypothesize an underpayment. The actual patterns in spending and RA payments remain an empirical question for the quantitative analyses in this study.

The key difference between Scenarios B and C is the moment consumers take into account their change in health status in choosing a health insurance plan. In Scenario B, consumers respond to the change in health status *after* year *t* (at the start of year *t+1*). This scenario could be expected for diseases that arise acutely (in year *t*), like heart failure or a cancer diagnosis. These are difficult to predict and cannot be anticipated when considering an insurance plan for the upcoming year. In this case, the expected underpayment for insurers that can be avoided through the disruption of the choice of insurance product by consumers is the payment gap shown in years *t+1* and *t+2* (and potentially the years after).

In Scenario C, consumers anticipate their imminent change in health status when choosing an insurance plan for year *t*, so before diagnosis. This scenario is more likely for predictable chronic diseases (e.g., due to accumulating health problems and examinations) where consumers might anticipate the onset of the chronic disease when selecting their new insurance plan, or when the eventual diagnosis is made late (the first reason why prospective RA may fall short to compensate these individuals, as explained in the Introduction). The avoidable underpayment for insurers in this scenario, shown by the payment gap in year *t*, is notably larger than in year *t+1*. Moreover, if the individual does not leave the insurer at the next opportunity, the insurer is faced with an additional year of a potentially avoidable, unprofitable contract. Therefore, the incentives for insurers to distort enrollment of the individuals newly diagnosed with a chronic disease, aggravated by the consumer response, are larger in Scenario C than in Scenario B.

The relevance of looking at RA outcomes and switching behavior of individuals who are newly diagnosed with a chronic disease can be illustrated by the following thought experiment. Imagine a scenario in which the average under/overcompensation for a group of individuals with a specific disease is zero. In that scenario, there are no selection incentives regarding this group as a whole. However, there can still be selection incentives regarding subsets of this group, for example, when newly diagnosed individuals are unprofitable to insurers and take into account their health change when switching insurance plans. In that scenario, despite the average fair compensation for spending of the disease group, insurers will prefer to deter the enrollment of newly diagnosed individuals.

This section demonstrates the relevance of two empirical questions in relation to the incentives for insurers to select against individuals with new chronic diseases: (1) to what extent do mean RA payments align with mean spending for this group, and (2) if and when do these individuals respond to their change in health status when choosing an insurance plan?

2.4 CONTEXT, DATA, AND METHODS

The approach taken for the analyses of this study comprises three steps: (1) identifying individuals who undergo a health shock in year *t* (group *g*), (2) tracking patterns in mean spending and mean RA payments for group *g* in the period of year *t–2* to *t+2*, and (3) tracking whether and when individuals of group *g* respond to the change in health status in the period of year *t–2* to *t+2* through insurer switching as an indicator. This section describes the context of the RA model used in our analyses, as well as the data and methods used.

2.4.1 Context and Data

The data for this study come from the Dutch basic insurance system, which is based on regulated competition among private insurers, with a mandatory insurance uptake for consumers. Regulatory aspects include an open enrollment policy for every new calendar year, a standardized benefits package, and premium rate restrictions. Aspects of competition include a free consumer choice of insurer and plan type in terms of deductible level, provider network, and coverage of out-of-network spending and instruments for insurers to enhance efficiency. The latter includes selective contracting of health care providers and freedom with respect to provider-payment methods. In 2017, the central year of our analysis, 15 insurers were active on the Dutch market, offering a wide selection of insurance plans (58 in total). While most insurers offered a variety of plans with different provider networks, out-of-network spending coverage, options for voluntary deductibles, and supplementary insurance plans, others exclusively specialized in cheaper plans. As a result, for individuals enrolled in a less generous plan with an insurer that offers a variety of plans, switching within the insurer is a viable option. Those enrolled in a less generous

plan from a single-plan insurer will have to switch insurers if broader coverage is desired, for instance, when becoming chronically ill.

Moreover, a prospective RA system[4] is in place to compensate insurers for predictable variation in health spending, thus mitigating incentives for risk selection. Recent studies have shown that the Dutch RA model largely compensates for predictable profits and losses, but not completely (Oskam et al., 2023; Withagen-Koster et al., 2023).
For the empirical analysis, two types of data are used: (1) administrative data used for RA and (2) diagnoses from electronic patient records. The first includes information on all individuals enrolled in a basic health insurance plan for the years 2015 until 2019 (N = 16.9 m–17.4 m). This administrative information was used for estimating the Dutch RA models for the years 2018 until 2022. Over these years, the model has undergone modifications that increased the total number of risk classes (dummy variables within the risk adjusters) from 193 to 226 (see Table A2.1 in the Appendix and Van Kleef et al. (2018a)).

In addition to the risk adjusters, the data include information on health spending. In the analysis, we follow the standard procedure of annualizing spending and weighing the outcomes with the fraction of the year the individual was enrolled in health insurance.[5] For the five consecutive years, the data show an upward trend both in terms of spending and in terms of the number of insured individuals. Mean spending in 2015 stood at €2,285, rising to €2,487 in 2019 (an 8.8% increase). However, for a fair comparison of the evolution of health spending and compensation over the included years, a correction factor is applied to spending in all years to account for inflation and other external effects. In this case, the year 2017 – with a mean spending of €2,355 – is taken as the benchmark. For the surrounding years, health spending has been adjusted by a factor equal to the relative difference in mean spending between that year and 2017.

The second type of data used for this study is diagnostic information from electronic patient records from GP practices, acquired from the Nivel Primary Care Database (Nivel-PCD) (Nivel, 2019; Vanhommerig et al., 2025). Specifically, this dataset includes 683 dummy variables (representing 683 diagnoses) that indicate the presence (1) or absence (0) of a health problem in a patient. Health problems are coded according to the International Classification of Primary Care by the Wonca International Classification Committee (Het Nederlands Huisartsen Genootschap, 2024). Out of the 683 diagnoses, 109 have been labeled as chronic diseases (Nielen et al., 2019). With 1.5 million individuals, these

4 The risk adjustment system includes separate models for somatic care, mental care, and out-of-pocket spending under the mandatory deductible. In this paper, we focus on the former, which covers roughly 90% of total spending under the basic health insurance.

5 Individuals may, for specific reasons, have canceled their health insurance at any point during the year (or have enrolled after January 1). Therefore, individual-level spending is annualized and weighted by the fraction of the year an individual was enrolled. For example, spending accumulated by individuals who were enrolled for half of the year is doubled and attributed a weight of 0.5 in the analyses.

data cover about 10% of the total Dutch population and are considered representative of the overall population for the corresponding years (2015–2019) (Van Kleef et al., 2020a). In addition, we generated a rebalancing weight for every individual in the Nivel-PCD using the raking method (Battaglia et al., 2009). The use of this weight guarantees that the prevalences and spending in the sample match those in the total population. In other words, it can be expected that, if the Nivel-PCD had covered the entire Dutch population, the findings would have been similar to those found in the weighted sample.

The data facilitate the identification of individuals who were diagnosed with a chronic disease within the period and the year of first registration of the diagnosis. We clustered specific diseases into overarching groups to indicate broader patient groups for our empirical assessment. Based on the available information, six subgroups are constructed: individuals who, in 2017, were diagnosed with (1) any chronic disease, (2) a chronic mobility disease, (3) diabetes, (4) a chronic heart disease, (5) a chronic lung disease, or (6) cancer.[6] Table A2.2 in the Appendix presents a complete overview of the respective clusters. Merging of the administrative data with the Nivel-PCD through unique individual-level identification codes enables the tracking of patterns in mean spending, mean RA payments, and insurer switching within the period *t–2* to *t+2* for individuals registered with a (new) chronic disease in year *t*.

2.4.2 Methods

The empirical analysis of this study comprised three steps. In the first step we identified individuals who are diagnosed with a new chronic disease in year *t* using the GP patient records within the Nivel-PCD. Through the multi-year combination of the data, the year in which a diagnosis is first registered was found. Moreover, a distinction was made between individuals who are diagnosed with a new disease in year *t* and were ill before, and those who had no chronic diagnosis before that point, effectively transitioning from healthy to chronically ill. Individuals were assigned to the group that was diagnosed with a (specific) chronic disease in year *t* if that disease was registered in year *t* (2017) and not in year *t–1* (2016). This distinction separates the six initial subgroups into twelve.

For the second step of the analysis, we tracked patterns in health spending and RA payments for the selected subgroups. The Dutch RA models of 2018 through 2022 were replicated using the relevant administrative information (for 2015 through 2019). The resulting individual-level RA compensations were then, on an aggregate level, compared to the mean health spending for the subgroups derived in step 1. The comparison demonstrates the mean financial profits or losses for insurers for the groups for the consecutive years.

6 Subgroups 3 through 6 are comprised of clustered diagnoses. Ambiguous diagnoses were excluded from the cluster to protect its validity.

Next, as a third step of this research, we tracked if and when consumers switched insurers in the period of year *t–2* to *t+2*. For the five contract years in the data (2015–2019), four measurable moments arise for individuals to adjust or switch their health insurance plan (e.g., before January 2016 for the contract period of 2016). As an indicator for the actual consumer response, we determine the percentage of individuals who switch insurers within the groups of interest for each switching moment. If this percentage peaks in a specific year, we consider that peak as an indication that people within the group respond to changes in health status through insurer switching. Ideally, we would have performed this analysis at the level of insurance plans and deductible level. Unfortunately, our data contains no information on plan membership and deductible level. Instead, we performed this analysis at the level of insurance companies. This is not a perfect representation of the complete consumer response; the consequences of this limitation are discussed in the Discussion (section 2.6).

In addition to detecting peaks in insurer switching within the five-year period, it is interesting to quantify the propensity of individuals within a group to switch insurers relative to the rest of the population. If the propensity to switch for a group is relatively high, the incentives for insurers to distort the choice of consumers in that group might be stronger, ceteris paribus (all other things being equal). However, since the onset of a chronic disease can be correlated with other consumer characteristics that may influence switching behavior (e.g., age), an additional analysis is required to determine the true influence of the onset of the chronic disease. Therefore, we applied a stepwise regression model to identify what characteristics (from the administrative data) are statistically significant predictors of switching between health insurers. Since age, sex, and the level of education have been proven to be influential factors in the propensity to switch insurers (see, for instance, van Vliet (2006) and Boonen et al. (2016), these variables were included. Additional information from the data is used (e.g., socioeconomic status) to find an association with insurer switching, while health-specific information is excluded.

The results from the stepwise regression indicate that age, sex, and level of education are indeed relevant factors. Females, younger individuals, and higher-educated individuals have a higher propensity to switch insurers than their peers. Moreover, the type of employment, being a student, socioeconomic status, and not living in a long-term care facility are significant predictors of switching between insurers. Table A2.3 in the Appendix provides information on the specific coefficients of these variables.

After identifying the controlling variables, a logit model was estimated to quantify the propensity to switch for the groups that were diagnosed with a (specific) chronic disease in year *t*. For the four separate windows of opportunity to switch in moment *y* = before January *t–1*, *t*, *t+1*, or *t+2*), the control variables are supplemented with a yes/no indicator for the respective disease/cluster diagnosed by the GP in year *t*. This process is repeated for all six

clusters defined in step 1, for all four moments, illustrating the effect of a given diagnosis on the probability $\pi_{i,y}$ of an individual *i* to switch insurer at moment *y*, through Equation 2.1:

$$\pi_{i,y} = \frac{exp(\beta_1 A_{i,y} + \beta_2 S_{i,y} + \beta_3 U_{i,y} + \beta_4 C_i + ...)}{1 + exp(\beta_1 A_{i,y} + \beta_2 S_{i,y} + \beta_3 U_{i,y} + \beta_4 C_i + ...)} \quad (2.1)$$

where *Ai,y*, *Si,y* and *Ui,y* represent the age, sex, and level of education, respectively, for individual *i* in moment *y*, and *Ci* indicates whether that individual was diagnosed with a (specific) chronic disease in year *t*. Crucially, *Ci* , holds four layers (I, II, III, and IV): individuals who are diagnosed with the disease in year *t* and were healthy before (group I), individuals who received a diagnosis in year *t* and already had a chronic disease (group II), individuals who already had the disease in year *t–1* and still do in year *t* (group III), and the rest of the population (group IV) as a reference group. The equation is completed by the other statistically significant predictors derived from the stepwise regression (the type of employment, being a student, socioeconomic status, and living in a long-term care facility), which are, for visual simplicity, omitted from the equation.

In summary, the analysis on switching behavior is split into two main steps: (1) checking for peaks in switching percentages *within* the considered groups to indicate (the moment of) a response to the change in health status and (2) running the logit model to quantify whether the individuals of a group switch more frequently than other groups (based on timeliness of diagnosis and versus the general population).

2.5 RESULTS

This section presents the results of the empirical analysis. First, the subdivision of the population into groups of individuals who are newly diagnosed with a chronic disease in year *t* is discussed. After, the trends of health spending, RA payments, and switching behavior for these groups are presented for the period *t–2* to *t+2*.

2.5.1 Identifying Individuals Who Undergo a Health Shock

Merging the results of the simulated RA models for the years *t–2* to *t+2* (2015–2019) with the respective data from the Nivel-PCD leads to a match for 1,548,441 individuals. To qualify, individuals need to be present, at a minimum, in the data for the years 2016 and 2017, recognizing the diagnosis of a new chronic disease in 2017. Table 2.1 shows the prevalence of the subgroups in the merged dataset. Based on the dataset, nearly 58% of insured individuals had at least one chronic disease in 2017, according to the GPs included in the Nivel-PCD. Here, common chronic diseases that are relatively manageable in terms of health impact make a significant contribution. For instance, 14.2% of the

Table 2.1. Prevalence of the subgroups based on chronic diseases identified in Nivel Primary Care Database

	Number of individuals (%)		
Total	1,548,441 (100%)		
At least one chronic disease	894,354 (58.8%)		
	Total	Previously healthy before diagnosis	Had a chronic disease before diagnosis
Diagnosed with any new chronic disease in year *t*	126,746 (8.2%)	39,013 (2.5%)	87,733 (5.7%)
Diagnosed with any chronic mobility disease in year *t*	15,897 (1.0%)	2,692 (0.17%)	13,205 (0.85%)
Diagnosed with diabetes in year *t*	4,403 (0.28%)	979 (0.063%)	3,424 (0.22%)
Diagnosed with heart disease in year *t*	16,199 (1.0%)	2,181 (0.14%)	14,018 (0.91%)
Diagnosed with lung disease in year *t*	3,427 (0.22%)	590 (0.038%)	2,837 (0.18%)
Diagnosed with cancer in year *t*	8,117 (0.52%)	1,579 (0.10%)	6,538 (0.42%)

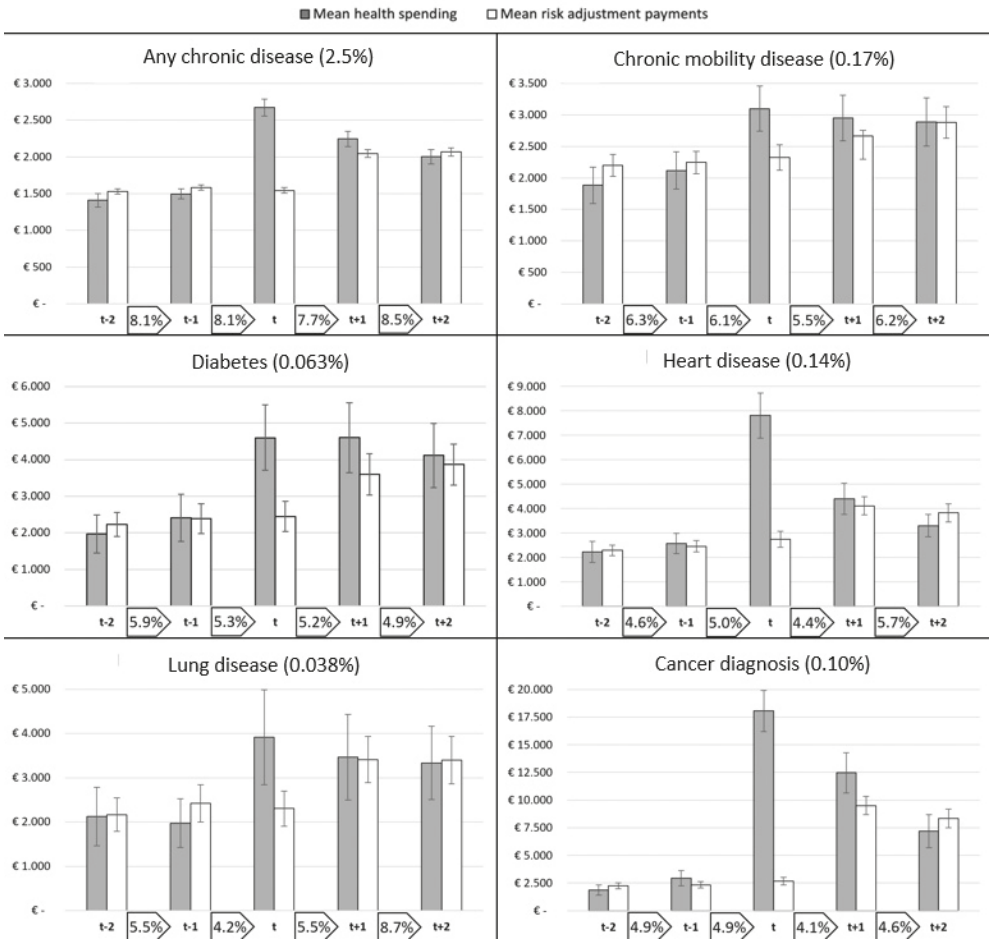
sample is diagnosed with hypertension without complications, 10.7% is diagnosed with dermatitis / atopic eczema, and 9.6% has asthma (not mutually exclusive).

Moreover, 126,746 individuals (8.2% of the sample) were diagnosed with a new chronic disease in 2017, of whom 39,013 (31%) were not diagnosed with any disease in 2016. Further observations include that 0.52% of the population was diagnosed with any cancer in 2017 (excluding skin cancer) and that only 13% (2,181 out of 16,199) of those diagnosed with a heart disease in 2017 were not diagnosed with any disease before.

2.5.2 The Evolution of Risk Adjustment Payments and Switching Around the Start of a Chronic Disease

For the identified subgroups of individuals who undergo a health shock, health spending is paired with RA compensations for the relevant years (2015–2019). In Figure 2.2, both are shown for six subgroups, exclusively covering the individuals who were healthy before the chronic disease was first registered in year *t* (i.e., groups under 'previously healthy before diagnosis' in Table 2.1). The disparity between the bars in the figure for mean spending and mean RA payments illustrates the mean annual payment gap (or surplus) to insurers for enrollees in the specific groups, like Figure 2.1. The bars include a 99% confidence interval to indicate the statistical significance of the pattern over the years and the difference between spending and compensation. Last, the percentages shown in the boxes between the years represent the percentage of enrollees within the selected group that switch insurer from one year to the next.

Figure 2.2. Patterns in spending, risk adjustment payments, and insurer switching for six subgroups of consumers who are diagnosed with a first chronic disease in year *t*.



Note: The six panels in the figure are not mutually exclusive, as individuals can be diagnosed with multiple diseases. The y-axes of the panels differ to facilitate a comparison over the years per disease. The percentage in brackets in the title of a diagram reflects the prevalence of the subgroup in the sample (as shown in Table 2.1). The lines superimposed on the bars denote 99% confidence intervals. The percentages between the labels on the horizontal axes represent the size of the group that switched insurers from 1 year to the next.[7] These percentages are separately shown in Figure A2.1 in the Appendix 8.

All subgroups start at similar levels of spending, but in the year of diagnosis, the six panels notably diverge. For individuals diagnosed with heart disease or cancer, both relatively acute indications, health spending surges first and then trends down later. For all diseases, prospective RA is, as expected, unable to accurately compensate for the

7 The overall switching rates in the population are 6.8%, 7.0%, 6.4%, and 7.3% for the respective moments.

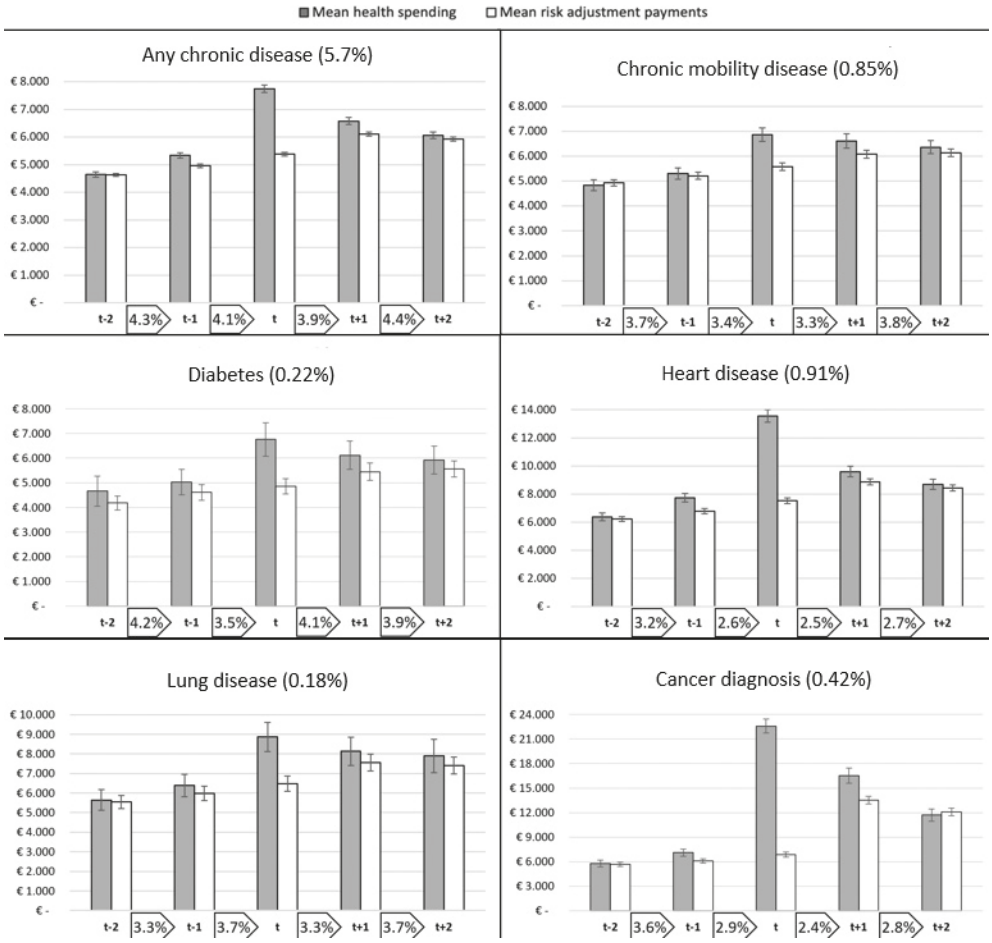
rise in spending in year *t*, but improves in the subsequent years. For all diagnoses, the confidence intervals between health spending and RA payments in year *t* do not overlap, indicating a statistically significant payment gap to insurers in the year of diagnosis. In the years before and after, the gap between spending and compensations is mostly no longer statistically significant. Moreover, in the year prior to diagnosis, a slight rise in health spending is observed for most subgroups, providing an indication of the latent health change or a belated diagnosis by the GP.

In terms of switching behavior from one year to the next, some interesting observations can be made. First, the rate of insurer switching is relatively high for the overarching group of individuals who were healthy before and were diagnosed with any chronic disease in year *t* (the top-left diagram).[8] However, there are no obvious peaks in switching for this group during the period of interest. For the five more specific subgroups, we find switching rates below that of the population (given in footnote 7) and some deviating patterns. Over the five-year period, individuals who were diagnosed with a chronic disease of the heart or lungs see an overall increase in the switching rate, while the opposite holds for the other three subgroups. Still, these numbers provide no convincing signal of a peak in insurer switching in response to the onset of the chronic disease. The exception to this finding is the peak in switching from year *t+1* to year *t+2* for individuals who are diagnosed with a lung disease in year *t*.

For the other six subgroups of our study, those that were diagnosed with a chronic disease on top of existing chronic unhealth, we find higher starting levels of spending in Figure 2.3. Nevertheless, we find similar patterns in the evolution of RA payments for these subgroups: statistically significant payment gaps in year *t*. However, these gaps linger in the first year after diagnosis. Only in the case of a cancer diagnosis, a surplus is shown in year *t+2*, albeit without statistical significance. In terms of switching, we find lower levels for all six subgroups than for the population (footnote t. For all groups, we find either a decrease in insurer switching or a stable pattern over the years.

8 The overarching group contains all individuals who were diagnosed with any of the 109 chronic diseases in 2017 and were healthy in 2016. The switching rates of this group exceed those of the general population (footnote 7) and the other subgroups (Figure 2.2). Additional research points out that the elevated switching rates come from relatively frequent, less severe chronic diseases such as asthma and eczema, which are more prevalent among younger individuals who tend to switch more frequently. The proven correlation between age and switching behavior helps to explain the result in Figure 2.2.

Figure 2.3. Patterns of spending, risk adjustment payments, and insurer switching for six subgroups of consumers who are diagnosed with a chronic disease in year *t* on top of an existing chronic disease.



Note: The six panels in the figure are not mutually exclusive, as individuals can be diagnosed with multiple diseases. The y-axes of the panels differ to facilitate a comparison over the years per disease. The percentage in brackets in the title of a diagram reflects the prevalence of the subgroup in the sample (as shown in Table 2.1). The lines superimposed on the bars denote 99% confidence intervals. The percentages between the labels on the horizontal axes represent the size of the group that switched insurers from 1 year to the next. These percentages are separately shown in Figure A2.2 in the Appendix.

The absolute differences in insurer switching between the subgroups and the overall population may be caused by underlying factors. Correcting for these individual-level characteristics, such as age and level of education (as discussed in the 'Methods' section) provides more insight into the propensity to switch in relation to the diagnosis in year *t*. Table 2.2 presents the propensity to switch insurers for all subgroups for the four moments to switch, derived from the logit model. For all six clusters, four groups are made (I, II, III, and IV): individuals who are diagnosed with the disease in year *t* and were healthy before (group I), individuals who received a diagnosis in year *t* and already had a chronic disease (group II), individuals who already had the disease in year *t–1* and still do in year *t* (group III), and the rest of the population (group IV) as a reference group.

Table 2.2. Propensity to switch insurer at four points in time, corrected for background characteristics, in contrast to the population

		From year t-2 to t-1	From year t-1 to t	From year t to t+1	From year t+1 to t+2
		Odds Ratio [99% C.I.]			
Any chronic disease	**I**	1.12 [1.06, 1.17]**	1.14 [1.08, 1.20]**	1.13 [1.07, 1.19]**	1.09 [1.04, 1.15]**
	II	0.96 [0.91, 1.01]	0.97 [0.93, 1.02]	0.98 [0.93, 1.03]	0.96 [0.92, 1.01]
	III	0.97 [0.95, 0.99]**	0.97 [0.95, 0.99]**	0.97 [0.95, 0.99]**	0.97 [0.95, 0.99]**
Chronic mobility disease	**I**	1.16 [0.94, 1.43]	1.20 [0.97, 1.48]*	1.16 [0.93, 1.44]	1.14 [0.93, 1.41]
	II	0.95 [0.84, 1.07]	0.95 [0.84, 1.08]	1.03 [0.91, 1.17]	1.02 [0.90, 1.15]
	III	0.92 [0.88, 0.96]**	0.92 [0.88, 0.97]**	0.97 [0.92, 1.02]	0.95 [0.90, 0.99]**
Diabetes	**I**	0.97 [0.68, 1.38]	0.90 [0.62, 1.30]	0.93 [0.64, 1.36]	0.77 [0.52, 1.13]
	II	0.96 [0.77, 1.20]	0.85 [0.67, 1.08]	1.08 [0.86, 1.35]	0.88 [0.70, 1.12]
	III	0.83 [0.79, 0.88]**	0.81 [0.76, 0.86]**	0.87 [0.82, 0.92]**	0.83 [0.78, 0.88]**
Heart disease	**I**	0.80 [0.61, 1.04]*	0.91 [0.70, 1.17]	0.84 [0.64, 1.10]	0.97 [0.76, 1.24]
	II	0.89 [0.78, 1.01]*	0.81 [0.71, 0.93]**	0.82 [0.71, 0.95]**	0.77 [0.67, 0.89]**
	III	0.90 [0.85, 0.95]**	0.89 [0.84, 0.95]**	0.86 [0.81, 0.92]**	0.89 [0.84, 0.94]**
Lung disease	**I**	0.94 [0.59, 1.51]	0.74 [0.43, 1.25]	1.02 [0.64, 1.64]	1.46 [0.99, 2.15]*
	II	0.79 [0.60, 1.04]*	0.99 [0.77, 1.28]	0.94 [0.72, 1.25]	0.92 [0.70, 1.20]
	III	0.83 [0.77, 0.90]**	0.86 [0.79, 0.93]**	0.84 [0.77, 0.92]**	0.84 [0.78, 0.91]**
Cancer	**I**	0.83 [0.61, 1.12]	0.88 [0.65, 1.19]	0.78 [0.56, 1.10]	0.74 [0.53, 1.04]*
	II	0.96 [0.80, 1.14]	0.87 [0.71, 1.05]*	0.74 [0.59, 0.93]**	0.75 [0.59, 0.94]**
	III	0.85 [0.79, 0.91]**	0.81 [0.75, 0.87]**	0.86 [0.80, 0.93]**	0.90 [0.84, 0.96]**

Note: Group I: Individuals who are diagnosed with the disease in year *t* and were healthy before. Group II: individuals who received a diagnosis in year *t* and already had a chronic disease. Group III: individuals who already had the disease in year *t-1* and still do in year *t*. The reference group in the model is composed of individuals who are excluded from the three respective groups (I, II, and III) per disease. The odds ratios shown result from a logit model that corrects for background characteristics, including age, sex, type of employment, and others. For the full overview, see Table A2.3 (Appendix). * for $p < 0.05$; ** for $p < 0.01$.

In Figure 2.2, higher switching rates were found for the overarching group of individuals who are diagnosed with a first chronic disease, relative to the general population. After correcting for background characteristics, we find that those individuals diagnosed with a chronic disease in year *t* who were healthy before are indeed more likely to switch insurer in all four moments ($\alpha = 0.01$). Individuals who already were diagnosed with a chronic disease and were diagnosed with an additional disease in year *t* (group II) do not deviate significantly from the population.

However, for our selection of specific (clusters of) diseases, we find slight deviations from the earlier patterns when correcting for the background characteristics. Individuals who were diagnosed with a chronic mobility disease and were healthy before are more likely to switch insurers before January of year *t* than the general population, although lower percentages were reported in Figure 2.2. Healthy individuals diagnosed with heart disease in year *t* have a lower propensity to switch before January *t–1*, and those diagnosed with such a disease in general have a lower likelihood to switch throughout the considered period. For individuals diagnosed with cancer, we find a lower propensity to switch before and after diagnosis. Only for those diagnosed with a chronic lung disease, while being healthy before, do we find a higher likelihood to switch insurer *after* diagnosis (before January *t+2*) (OR: 1.46). In general, we find individuals who already were chronically ill to be less likely to switch insurers than the overall population. Except for the group diagnosed with a chronic mobility disease, we find no evidence of an increase in insurer switching behavior in anticipation of, or in response to, the diagnosis of a specific, or clustered, chronic disease. Also, at a 99% confidence level, it cannot be claimed that for any of the subgroups shown here in any of the switching moments, individuals are more likely to change insurer than in the prior year (because of the structural overlap in confidence intervals between the considered moments). In summary, for the (clusters of) diseases we selected, we find no elevated switching rates in the years surrounding diagnosis (except for the chronic mobility disease group before year *t* and the lung disease group before year *t+2*), while we do find a significant effect for the group of individuals who are diagnosed with a first chronic disease in general. This implies that for other diseases not included in our clusters, elevated switching rates should be found. Additional analysis confirms the presence of above-average switching rates for individuals diagnosed with relatively less severe and more common chronic conditions like asthma and eczema.

2.6 DISCUSSION

In this research, we tracked the five-year evolution of health spending, compensations to insurers through RA, and insurer-switching behavior for individuals who are diagnosed with a new chronic disease. The dynamics between the three provide an indication of the incentives for insurers to distort the natural enrollment of individuals who are diagnosed

with a chronic disease. Using multi-year GP data from the Nivel-PCD, we identified subgroups of the Dutch population who were diagnosed with a chronic disease in 2017 and found all subgroups to be substantially underpaid by the prospective RA model in place in the year of diagnosis: a logical consequence of the belated recognition of the onset of a chronic disease in prospective RA. In the subsequent years, RA compensations increase but often do not completely cover health spending.

We argued that the incentives for insurers to engage in risk selection toward individuals who are first diagnosed with a (new) chronic disease depend on the presence of a payment gap and whether and when these individuals anticipate the health shock in their plan selection. Through examination of the insurer switching behavior of individuals – as an indicator of consumer response – we find no obvious peaks in switching in years before or after the chronic disease was first diagnosed. However, we do find that – on average – people first diagnosed with a chronic disease are more likely to switch insurers than the rest of the population, both before and after the first diagnosis. For our selection of specific diseases, however, we do not find strong indications that newly diagnosed individuals act in anticipation of, or in response to, their change in health status.

Importantly, the findings with respect to consumer response are subject to caveats. As mentioned in the 'Methods' section, switching between insurers as our only indicator for responsive behavior is incomplete. In response to an imminent health change, a logical action by individuals could be to adjust their voluntary deductible or switch to a broader insurance plan offered by the same insurer. Therefore, the insurer switching rates used in our analyses are likely underestimations of actual switching behavior. That might be a reason why our findings do not directly match earlier studies on the association between health plan enrollment and health status. Another reason could be that the level of basic coverage in the Netherlands is quite substantial as a baseline, or that consumers may well be enrolled in a relatively generous plan already. Both reduce the need for a change of insurance plan in response to a diagnosis. Despite the standardized basic benefit package in the Netherlands, insurance products may differ on the level of contracted providers. Packages with unrestricted access to providers are typically more expensive. A more thorough analysis of the plan package design and deductible choice of individuals in this context could find results that align more with earlier studies. Moreover, a related element affecting our findings is the fact that the results depend on the Dutch context and the available health plans on the market during the specific years included in the research. Within that context, the design of health plans can be the consequence of actions of selection by both insurers and consumers. Therefore, switching behavior and plan design share an endogenous relationship. This dynamic should be considered in further research, for instance, by incorporating the plan attributes before and after switching.

Another limitation of our analysis is the focus on a subset of chronic diseases. An interesting direction for future research is to track spending, RA payments, and switching for other diseases. If specific diseases are in fact problematic in the light of selection incentives (e.g., a payment gap and an active anticipation of the health change by consumers), regulators could consider concurrent RA for that disease (e.g., diagnoses from year *t* are used to 'predict' health spending in year *t*) to reduce the potential incentives for risk selection by insurers. However, concurrent RA is not without flaws either: the endogenous link between health care consumption and financial compensation in the same year may reduce incentives for insurers to control costs (Van de Ven & Ellis, 2000a). Generally, regulators could think of a hybrid system with prospective risk adjusters for diseases that are unlikely to be anticipated by consumers and concurrent risk adjusters for diseases that can be anticipated to reduce incentives for risk selection.

In sum, by tracking spending and RA payments over the period *t–2* to *t+2* for individuals diagnosed with a new chronic disease in year *t*, we find a substantial payment gap in year *t* and, to a lesser extent, in prior and/or subsequent years. Our analysis of insurer switching shows that – on average – people first diagnosed with a chronic disease are more likely to switch insurers than others. This could indicate that consumers respond to their change in health status, which could aggravate the selection incentives for insurers that result from the observed payment gaps. Circling back to the central question of this study: for diseases that are statistically significantly undercompensated by the RA model in year *t+1* and/or *t+2*, we argue that the model does not compensate adequately enough. Competing health insurers can potentially steer the enrollment behavior of individuals diagnosed with such unprofitable diseases when that information becomes known to insurers. If these individuals are concentrated at specific insurers due to enrollment in specific health plans, no level playing field for insurers may exist. To assess the extent to which undercompensations in year *t* are truly problematic, a deeper analysis of switching and enrollment behavior is required, in addition to studying selection strategies by insurers. Further research is therefore needed to analyze switching patterns at the level of plan attributes (e.g., deductible level and provider network) and repeat our analysis for (other) diseases of interest.

2.7 APPENDIX

Table A2.1. Risk adjusters in the Dutch risk adjustment models of the years 2018 until 2022 and information on spending for the respective years (2015 until 2019) underlying the models

	Number of risk classes (dummy variables)				
	2018	2019	2020	2021	2022
Risk adjusters					
Age interacted with sex	42	42	42	42	42
Pharmacy-based cost groups	34	38	38	39	43
Diagnosis-based cost groups				27	27
Primary diagnosis-based cost groups	16	16	16		
Secondary Diagnosis-based Cost Groups	8	8	8		
Durable medical equipment cost groups	11	11	11	15	15
Source of income/education interacted with age	25	25	36	36	36
Zip code clusters for region	10	10	10	10	10
Socioeconomic status	12	12	12	12	12
Household size interacted with age	13	13	13	13	13
Multiple-year high-cost groups	9	9	9	9	9
Physiotherapy-diagnosis cost groups	5	5	5	5	5
Home care spending in the previous year	8	10	10	10	10
Multiple-year pharmaceutical costs *and* historical somatic morbidity					4
Total	**193**	**199**	**210**	**218**	**226**
Information on spending and age	2015	2016	2017	2018	2019
Mean spending in the population	€2,285	€2,352	€2,355	€2,408	€2,487
Mean spending after correction	€2,355	€2,355	€2,355	€2,355	€2,355
Number of insured individuals	16,945,802	17,036,618	17,128,725	17,256,295	17,365,571
Number of insured years	16,672,147	16,749,830	16,839,948	16,951,700	17,041,789
Under the age of 18	20.4%	20.2%	19.9%	19.7%	19.5%
18 to 34 years of age	20.4%	20.4%	20.6%	20.9%	21.0%
35 to 54 years of age	27.9%	27.5%	27.1%	26.7%	26.3%
55 to 69 years of age	19.3%	19.4%	19.4%	19.4%	19.5%
70 to 79 years of age	7.7%	8.0%	8.4%	8.8%	9.1%
Aged 80 or above	4.4%	4.5%	4.6%	4.6%	4.8%

Table A2.2. Overview of the clustered diseases for the defined subgroups

Cluster	Diseases included
Mobility	Rheumatoid arthritis and allied conditions (L88), Osteoarthrosis of hip (L89), Osteoarthrosis of knee (L90), Osteoarthrosis, other (L91), Osteoporosis (L95).
Lung	Chronic bronchitis/ bronchiectasis (R91), Emphysema/chronic obstructive pulmonary disease (R95).
Heart	Angina pectoris (K74), Other and chronic ischaemic heart disease (K76), Heart failure (K77), Pulmonary heart disease (K82), Hypertension complicated (K87), Stroke/cerebrovascular accident (K90), Atherosclerosis (K91), Other arterial obstruction/peripheral vascular disease (K92).
Cancer	Malignancy, not otherwise specified (A79), Hodgkin's disease/lymphoma (B72), Leukaemia (B73), Malignant neoplasm blood, other (B74), Malignant neoplasm stomach (D74), Malignant neoplasm colon/rectum (D75), Malignant neoplasm pancreas (D76), Malignant neoplasm digestive, otherwise/not otherwise specified (D77), Malignant neoplasm nervous system (N74), Malignant neoplasm bronchus/lung (R84), Malignant neoplasm respiratory, other (R85), Malignant neoplasm thyroid (T71), Malignant neoplasm kidney (U75), Malignant neoplasm bladder (U76), Malignant neoplasm urinary, other (U77), Malignant neoplasm cervix (X75), Malignant neoplasm breast female (X76), Malignant neoplasm female genital, other (X77), Malignant neoplasm prostate (Y77), Malignant neoplasm male genital, other (Y78).

Table A2.3. Odds Ratios for the background characteristics on insurer switching for the four switching moments considered, resulting from the stepwise regression

		From year *t-2* to *t-1*	From year *t-1* to *t*	From year *t* to *t+1*	From year *t+1* to *t+2*
Variable	Description	Odds Ratio			
Sex	Man	0.945	0.93	0.913	0.907
	Woman	1	1	1	1
Age	Under 18	4.968	5.16	5.815	6.745
	From 18 to 34	6.726	7.129	7.513	8.137
	From 35 to 54	4.199	4.481	4.764	5.558
	From 55 to 69	2.877	2.966	3.083	3.672
	From 70 to 79	1.506	1.61	1.668	1.844
	80 or above	1	1	1	1
Source of income	Unable to work, unemployed or government allowance.	0.828	0.874	0.928	0.856
	Student	1.091	1.089	1.094	1.056
	Self-employed	1.132	1.108	1.168	1.162
	High educated	1.552	1.468	1.631	1.646
	Reference group	1	1	1	1
Socioeconomic status	Very low	1.036	1.183	1.125	1.028
	Low	1.035	1.144	1.078	1.037
	Mid	1.051	1.105	1.093	1.085
	High	1	1	1	1
Living in long-term care facility	Yes	0.545	0.595	0.497	0.519
	No	1	1	1	1

Figure A2.1. Switching percentage over four windows for individuals who are diagnosed with a chronic disease in year *t* and who were healthy before

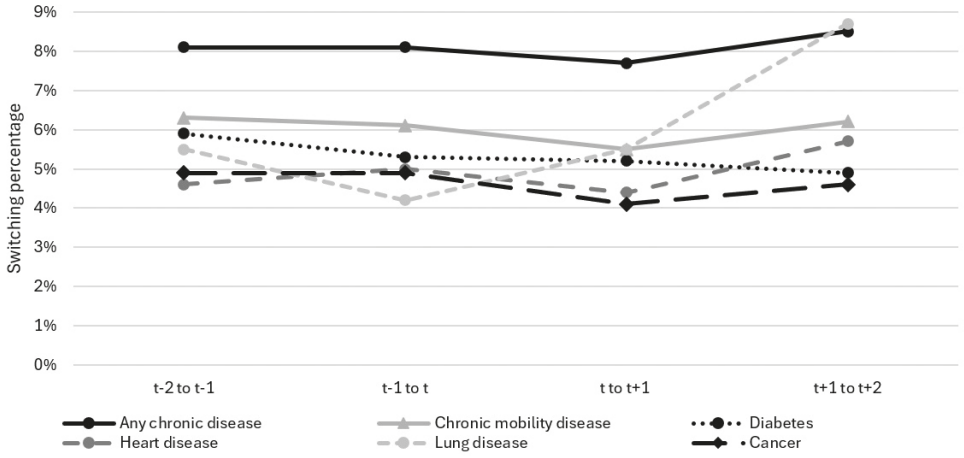
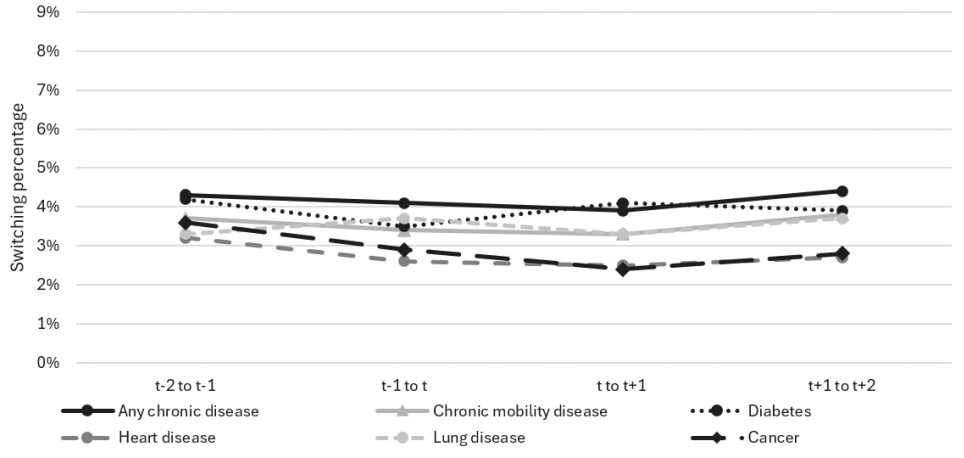


Figure A2.2. Switching percentage over four windows for individuals who are diagnosed with a chronic disease in year *t* and who were already diagnosed with a chronic disease before



3

SELECTION INCENTIVES THAT RESULT FROM HETEROSCEDASTICITY OF RESIDUAL SPENDING

Based on: Oskam, M., Van Kleef, R. C., & Douven, R. (2024). Heteroscedasticity of Residual Spending After Risk Equalization: a Potential Source of Selection Incentives in Health Insurance Markets with Premium Regulation. *The European Journal of Health Economics, 25*(3), 379-396. https://doi.org/10.1007/s10198-023-01592-9

Abstract

Many community-rated health insurance markets include risk equalization (also known as risk adjustment) to mitigate risk selection incentives for competing insurers. Empirical evaluations of risk equalization typically quantify selection incentives through predictable profits and losses net of risk equalization for various groups of consumers (e.g., the healthy versus the chronically ill). The underlying assumption is that absence of predictable profits and losses implies absence of selection incentives. This paper questions this assumption. We show that even when risk equalization perfectly compensates insurers for predictable differences in mean spending between groups, selection incentives are likely to remain. The reason is that the uncertainty about residual spending (i.e., spending net of risk equalization) differs across groups, e.g., the risk of substantial losses is larger for the chronically ill than for the healthy. In a risk-rated market, insurers are likely to charge a higher profit mark-up (to cover uncertainty in residual spending) and a higher safety mark-up (to cover the risk of large losses) to chronically ill than to healthy individuals. When such differentiation is not allowed, insurers face incentives to select in favor of the healthy. Although the exact size of these selection incentives depends on contextual factors, our empirical simulations indicate they can be nontrivial. Our findings suggest that – in addition to the equalization of differences in mean spending between the healthy and the chronically ill – policy measures might be needed to diminish (or compensate insurers for) heteroscedasticity of residual spending across groups.

3.1 INTRODUCTION

Contrary to other insurance products, health insurance is often subject to some form of premium regulation to protect the affordability of coverage for high-risk people. While car insurers apply the perceived risk of damage to adjust the individual premiums, for instance, through age or historical claims data, health insurers are typically obliged to apply community-rating per insurance product. As a result, consumers are charged the same premium for the same product – irrespective of their risk of medical expenditure –, confronting insurers with predictable profits on healthy enrollees and predictable losses on unhealthy enrollees, subsequently resulting in risk selection incentives (Newhouse, 1996; Van de Ven & Schut, 2011).

To mitigate selection incentives, regulated health insurance markets generally include a system of risk equalization (also known as risk adjustment) to compensate insurers for predictable spending variation using individual characteristics (risk adjusters) such as age, sex, (prior) diagnoses, and the (prior) use of specific pharmaceuticals. Through incremental advancements made over the past decades, risk equalization models have become increasingly accurate in compensating for predictable spending variation and reducing predictable profits and losses for subgroups of enrollees.

Studies on the design and evaluation of risk equalization models typically assume (either explicitly or implicitly) that – in the absence of predictable profits and losses – there is no financial incentive for insurers to distort the natural enrollment of individuals. As a result, insurers would need to resort to efficient contracting of care services to generate profit, improving the overall efficiency of health care delivery and the functioning of the insurance market (Enthoven, 1988a; Glazer & McGuire, 2000; Van Kleef et al., 2013a; Van Kleef & Van Vliet, 2022b; Van de Ven & Ellis, 2000a; Van de Ven & Schut, 2011). Therefore, over the past three decades, research and policy implementations regarding risk equalization design have focused on reducing the predictable profits and losses to diminish the incentives for risk selection (for an overview, see Ellis et al. (2018)). Although the substantial advancements made to risk equalization models did result in notable improvements in predictive strength, the most sophisticated morbidity-based models currently in place do not eliminate predictable profits and losses (McGuire et al., 2020; Van Kleef & Van Vliet, 2022b).

Through this paper we make, to the best of our knowledge, a new contribution to the academic literature. We argue that even when risk equalization would perfectly compensate for the predictable profits and losses of identifiable subgroups of individuals, incentives for risk selection remain. In other words, the absence of a predictable profit/loss for risk type g does not imply the absence of selection incentives towards risk type g. Our argument is that the distribution of residual spending (i.e., spending net of risk equalization payments) is likely to vary considerably across risk types. More specifically,

the variance of residual spending is likely to increase with expected health spending, exposing insurers to more uncertainty and a greater risk of substantial losses from risk types with high expected spending (e.g., those with a pre-existing condition) than from healthy risk types, ceteris paribus. In the conceptual framework of this paper, we explain and demonstrate how 'heteroscedasticity' of residual spending can be a source of risk selection incentives: when insurers are not allowed to risk-rate their premiums, they are likely to prefer risk type with low variance in residual spending over risk types with high variance in residual spending, ceteris paribus. Hence, perfect equalization of the *mean* expected result for different risk types is no guarantee for the absence of selection incentives towards these risk types. Therefore, the primary objective of this paper is to quantify the heteroscedasticity of residual spending. The precise size of the effects will depend on the risk equalization system of an insurance system and corresponding (inter)national regulations. In this paper, therefore, we use the Netherlands as a case study and simulate the selection incentives using Dutch administrative data.

The paper is structured as follows. Section 'Conceptual framework' provides a conceptual framework on how heteroscedasticity of residual spending may lead to selection incentives using 1) literature on risk management in financial markets and 2) capital requirements imposed for insurers by regulatory authorities. In the section 'Data and methods' the data and methods used for our simulation analysis are described. In an explorative analysis we compare the standard deviation of residual spending across risk classes to indicate differences in financial uncertainty and compare the 99.5th percentile of residual spending across risk classes to approximate differences in the risk of substantial losses. In the section 'Results', we simulate the effects of such heteroscedasticity of residual spending on selection incentives. In the section 'Size of the problem and potential solutions', we discuss the importance of these selection incentives and potential strategies to correct for these incentives. Finally, in the section 'Discussion' our main findings and their implications are summarized and discussed.

3.2 CONCEPTUAL FRAMEWORK

The problem can be illustrated through the following thought experiment: suppose insurers can select against or in favor of the following two types of consumers: high-risks (H) with a pre-existing condition and low-risks (L) without a pre-existing condition. From the viewpoint of insurers, mean per person expected spending equals €1,000 for L and €5,000 for H. Both types comprise 50% of the population, setting the overall average to €3,000. Assume insurers operate in a market with community-rated premiums (per insurance plan) and a sophisticated risk equalization model that perfectly compensates insurers for the difference in *mean* expected spending between L and H. More specifically, insurers receive a risk equalization payment of €2,000 (€5,000–€3,000) for the above-

average expected spending of H and contribute a risk equalization payment of €2,000 (€1,000–€3,000) for the below-average spending of L. From the insurers' perspective, the mean per person expected spending *net of* risk equalization equals €3,000 for both risk types (which would be covered by the community-rated premium). So, the insurers' *expected* financial result for L is equal to that for H. Nevertheless – and that is the key point of this paper – the uncertainty surrounding the expected financial result is likely to be higher for H than for L. The reason is that the standard deviation of the *actual* financial result per individual increases with the level of expected spending, as will be shown later in this paper. Based on two arguments (discussed in section 'Variation in uncertainty of financial return' and section 'Variation in solvency requirements', respectively), we hypothesize that – under the requirement of community-rating – this heteroscedasticity leads insurers to prefer enrollment of L over H, despite the fact that insurers are perfectly compensated for the difference in mean spending between the two risk types.

3.2.1 Variation in uncertainty of financial return

A first argument why heteroscedasticity of residual spending results in selection incentives stems from the assumption that insurers are likely to prefer a relatively *certain* financial return over a relatively *uncertain* return of the same size. Derived from examples of risk-averse behavior by insurers, such as the uptake of reinsurance or the use of loading factors based on group size, literature suggests that complete risk neutrality of insurers is unlikely (Eeckhoudt & Schlesigner, 2013; Layton et al., 2016). Simplifying the business of insurance, providing coverage of costs for a group of consumers can be considered a financial investment where the insurer invests capital for the operation of insurance and expects a return. Following theory of investment risk, the investor (insurer) is likely to desire a higher *expected* return on his investment if the *actual* return is more uncertain (Whitbeck & Kisor, 1963). This implies that – if insurers in our example were allowed to charge risk-rated premiums – they would have been inclined to charge a higher loading fee – or more specifically: a higher *profit mark-up* – to H than to L, ceteris paribus. This assumption corresponds with the work on insurance risk premiums by Kahane (1979), stating that the expected return on any investment portfolio incorporates a risk loading fee, proportional to the standard deviation of the portfolio.

Contrary to unregulated markets with risk-rated premiums, regulated markets with community-rated premiums prohibit insurers from charging different profit mark-ups to different groups. In community-rated markets, the variation in profit mark-ups that would have occurred in a risk-rated market serves as an approximation of the selection incentives. Although the exact size of these selection incentives is unknown, metrics from investment risk management can be used to understand the link between the degree of uncertainty and a corresponding desired excess return. When considering multiple stocks for investment, traditional measures to quantify the extent to which one investment op-

tion is preferred over another include the coefficient of variation,[1] the Beta coefficient[2] and the Sharpe ratio (Equation 3.1) (Brief & Owen, 1969; Sharpe, 1994; Rosenberg & Guy, 1976). While the three measures vary in their specification, they all rely on the standard deviation of (historical) results to quantify the riskiness of investment options. An important beneficial aspect of the Sharpe ratio over the other two measures is its potential to derive the desired excess return on investment for enduring risk. In other words, the Sharpe ratio helps to indicate the desired profit mark-up for specific levels of uncertainty. For investment x, the Sharpe ratio (*S*) is calculated as follows:

$$S_x = \frac{(r_x - R_f)}{Standard\ Deviation(r_x)} \qquad (3.1)$$

Here, r_x represents the average return on investment x and *Rf* stands for the best available rate of return of a risk free asset. For investors deciding between assets for investment, the asset with the largest Sharpe ratio is preferred (Sharpe, 1994). Alternatively, by setting a desired Sharpe ratio, the corresponding required value of r_x (= excess return, or 'price of uncertainty') can be calculated for any combination of Sharpe ratio and endured standard deviation. Here, the Sharpe ratio reflects a stance of profit-seeking as opposed to the endured risk. We will elaborate on this in the Methods section.

3.2.2 Variation in solvency requirements

The second argument why heteroscedasticity of residual spending may generate selection incentives lies in the solvency regulations typically instructed to insurers. For example, European Union (EU) legislation requires health insurers to have sufficient financial capital to remain solvent (over the period of one year) with a certainty of 99.5%, or, conversely, to cover a 1-in-200-year catastrophic financial shock (EUR-Lex, 2009).[3] The directive has been implemented to protect insurance firms (and their customers) from financial disasters based on their respective risk profiles and capital reserves. Given the difference in variation of residual spending between L and H in our example, the potential financial loss incurred in a 1/200 chance risk of ruin could be significantly higher for groups of H-type consumers than for L-types. Moreover, the magnitude of financial losses above any chosen threshold could be greater for H than for L, ceteris paribus. Consequently, the capital requirements should be higher when enrolling H-type consumers than when enrolling L-type consumers. If insurers were allowed to charge risk-rated

1 Coefficient of Variation (CV_x): $\frac{\sigma}{\mu}$. The coefficient of variation is a measure used to indicate the risk/return trade-off by dividing the standard deviation of the return by the mean (expected) return. A lower value for CV indicates a better risk/return trade-off.

2 Beta coefficient (β_x): $\frac{Covariance(r_x, R_m)}{Variance(R_m)}$. The Beta coefficient is a measure to value the relative risk of an asset versus the overall market. r_x depicts the return on a considered stock, while R_m depicts the return on the overall market. The standard deviation – i.e., the square root of the variance – directly influences the *Beta*, serving, to some extent, as a proxy for risk.

3 Articles 101 and 104 of the European Solvency II Directive (2009/138/EC) dictate a solvency capital requirement to insurers at a 99.5% confidence level.

premiums, there would be an inclination to charge a higher loading fee – or more specifically: a higher *safety mark-up* – to H than to L, ceteris paribus (Olivieri & Pitacco, 2015). Since insurers in our example operate in a regulated market with enforced community-rating (and thus are *not* allowed to risk-rate premiums), the different respective risks of ex-post losses lead insurers to prefer enrollment of L over H. To quantify these incentives, researchers can calculate the difference in solvency requirements when enrolling H versus L. For example, to indicate the risk corresponding to a 1/200 outlier in residual spending, the 99.5th percentile of the distribution of mean residual spending can be calculated for a given portfolio size. By doing this separately for H and L, researchers can approximate the difference in capital requirements for a portfolio of H-type enrollees versus a portfolio of L-type enrollees. Consequently, the 'cost' associated with this difference in capital requirements can serve as a proxy for selection incentives under community-rated premiums.

In practice, the EU solvency legislation encompasses more than the 1/200 risk in residual spending. Specifically, in the Commission Delegated Regulation (EUR-Lex, 2015a),[4] supplementing the Solvency II Directive (EUR-Lex, 2009), the 'health underwriting risk module' is introduced as one explicit section of the diversified Solvency Capital Requirement (SCR)[5] for health insurers. The module covers the premium and reserve risk borne by insurers, dependent of the respective composition of the portfolio of enrollees of an insurer. The contribution to the capital requirement by the health module, from here onwards *SCR*, depends on the following three variables: the total number of enrollees in the portfolio of an insurer, the level of financial reserves, and the annual income through premiums and risk equalization payments. Since the risk equalization payments vary across risk types, the potential enrollment of L versus H, or a group of either type, alters the amount of capital legally required.

Therefore, to comply with the SCR (further operationalized in the Methods section) in a risk-rated market, different safety mark-ups would be applied for the two types of consumers. These theoretical safety mark-ups for the two types may be found through a subsequent analysis on the difference between the respective SCRs and the associated opportunity costs. Moreover, to comply with the solvency requirements, the extra amount of required capital cannot be freely invested elsewhere and will have to be frozen. Potentially, relatively safe investments such as government bonds could be allowed but will likely earn the investor fewer returns than unrestricted investments. Alternatively, if an insurer would have to loan the extra capital to comply with the SCR, interest will have to be paid. The cost of capital for the difference in SCR between the two types of consum-

4 Articles 144–149 of the Commission Delegated Regulation (EU) 2015/35 supplementing Directive 2009/138/EC define the capital requirement for health insurers subject to a defined variance parameter and the volume of an individual insurer.

5 The health underwriting risk module is one part of the complete Capital Requirements for insurers. Other risk modules are disregarded in this paper for reasons of simplification. As a result, the Capital Requirements discussed in this paper are potentially underestimated.

ers therefore serves as an approximation of the selection incentives between these types. Say, for instance, that the per-person capital requirements are €400 higher for consumer type H than for consumer type L. An interest rate of 10% on a loan would then imply that the cost of meeting the capital requirement is €40 higher for enrolling an H-type consumer than for enrolling an L-type counterpart. If such costs cannot be reflected in risk-rated safety mark-ups, because of a community-rating mandate, this disparity generates selection incentives for insurers.

3.2.3 Group size and the law of large numbers

As mentioned above, the group size of individuals pooled is a crucial factor in determining the level of uncertainty borne by an insurer. The law of large numbers reduces both the standard deviation of the mean (residual) spending and the risk of outliers in mean (residual) spending (Layton et al., 2016). Hence, the selection incentives that result from both the 'uncertainty of financial return' and the 'risk of ruin' are also a function of the group size. More specifically, the absolute differences in 'uncertainty of financial return' and 'risk of ruin' between risk groups are likely to decrease with the size of these groups, as we will discuss and show in our empirical analyses below. This implies that selection incentives might be weaker, as insurers can attract larger numbers of individuals from these respected groups.

3.3 DATA AND METHODS

The goals of our analyses are 1) to quantify the heteroscedasticity of residual spending across selective subgroups derived from risk adjusters of the equalization model (from here onwards referred to as 'risk groups') and 2) to approximate its potential effect on selection incentives. This section describes the data and the methods used to achieve these goals. In order to disentangle 'heteroscedasticity of residual spending' and 'predictable profits and losses' (i.e., the traditional approach of measuring selection incentives), our analyses are focused on risk groups for which mean residual spending equals zero (which is typically the case for groups that are explicitly flagged by risk adjusters in the risk equalization model, assuming payment weights for these risk adjusters are estimated through the ordinary least squares method).

3.3.1 Data

The data used for this research comes from the Dutch basic health insurance and was originally used to calibrate the Dutch risk equalization model of 2021. The data include individual-level medical spending covered under the basic health insurance and risk adjuster information for nearly the entire Dutch population in 2018. The risk equalization system includes separate models for somatic care and mental care. In this paper we focus on the former, which covers about 90 percent of total spending under the basic health in-

surance. The information used to predict individual-level spending contains eleven types of risk adjusters that collectively account for more than 200 risk classes that take the form of dummy variables with a value of 1 (0) for individuals who are (not) a member of that class. The 11 types are as follows: age interacted with sex, institutional status interacted with age, clusters of zip codes based on regional factors, socioeconomic status interacted with age, household size interacted with age, pharmacy-based cost groups (PCGs), diagnosis-based cost groups (DCGs), durable medical equipment cost groups (DMECGs), multiple-year high-cost groups (MYHCs), physiotherapy diagnosis cost groups (PDCGs), and prior spending on home care (PSHC) (Van Kleef et al., 2018a). For the purpose of this paper, we classify the PCGs, DCGs, DMECGs, MYHCs, PDCGs and PSHC as 'morbidity adjusters', as they are directly derived from (prior) utilization of health care. In all analyses, we follow the standard procedure of annualizing spending and weighting the outcomes with the fraction of the year the individual was enrolled in health insurance.[6]

3.3.2 Quantifying the heteroscedasticity of residual spending

To obtain insight into the heteroscedasticity of residual spending from the Dutch risk equalization model, we first replicate the model that was actually in place in 2021.[7] Next, we identify risk groups defined by risk adjusters, such as age or morbidity status. Since these groups are explicitly flagged by risk indicators in the risk equalization model, the mean residual spending of these groups equals zero (a property of the ordinary least squares method that is used here to derive payment weights). In a third step, to quantify the heteroscedasticity of residual spending, we estimate the standard deviation (as a proxy for the uncertainty in residual spending) and the 99.5th percentile of residual spending (as a proxy for the risk of ruin) for each risk group. Variation in these measures across risk groups provides an indication of the heteroscedasticity in residual spending across risk types.

Moreover, given the relevance of group size through the law of large numbers, we also simulate the standard deviation and 99.5th percentile of the *mean* residual spending for different portfolio sizes. For each risk group *g* and a given portfolio size *N*, we run 1,000 simulations in which we randomly select *N* consumers from risk group *g* for which we

6 Individuals may, for particular reasons, have cancelled their health insurance at any point during the year (or have enrolled after January 1st). In all our analyses, individual-level expenditures are annualized and weighted by the fraction of the year an individual was enrolled. For example, expenditures incurred by individuals that were enrolled for half of the year is doubled and attributed a weight of 0.5 in the analyses. In the subsequent parts of this study, the prevalence of any group is consistently measured in insured years, rather than a total number of contracts.

7 For simplification we add a risk adjuster that contains two classes: 'consumers that qualify for at least one of the six morbidity adjusters' and 'consumers that do not qualify for morbidity adjusters at all'. These could be framed as 'any pre-existing condition' or 'no pre-existing condition', respectively, in terms of the six morbidity indicators. Our motivation for adding this risk adjuster is that the groups with/without a pre-existing condition ('healthy' versus 'unhealthy') play an important role in our analysis of heteroscedasticity of residual spending. Inclusion of this risk adjuster ensures that the mean residual spending equals zero for both groups, simplifying the presentation and interpretation of results.

calculate mean residual spending. For these 1,000 values we calculate the standard deviation and 99.5th percentile. An increase in portfolio size is expected to result in a decrease of the two measures. We run our simulations for four portfolio sizes: 1,000, 10,000, 100,000, and 1,000,000 individuals.

3.3.3 Quantifying the selection incentives generated by heteroscedasticity of residual spending

The standard deviation and 99.5th percentile of (mean) residual spending do, as standalone measures, not directly provide an indication of the *size* of selection incentives on a *community-rated market*. To obtain this indication, we simulate the profit and safety mark-ups that insurers would – most likely – have charged to the various types of consumers in a *risk-rated market*.

First, we derive the profit mark-up for risk group *g*: the mark-up that an insurer would charge, based on the uncertainty of the financial return on group *g*, to each individual of that group. As shown in section 'Variation in uncertainty of financial return', the desired excess return on an investment, r_x, for enduring more uncertainty on its result can be found by selecting a Sharpe ratio, *S*. Typically in equity markets, a Sharpe ratio of 0.2 is found but levels over 2.0 can be achieved for well-returning assets as well as negative ratios for poor performances (Bayraktar et al., 2009; Hodges et al., 1997; Wang et al., 2022). Taking a conservative ratio of 0.2 would be in conformance with typical investment markets. Alternatively, the ratio could be altered to reflect stronger or weaker profit-seeking behavior. It could be argued that not-for-profit semi-institutional insurance markets, such as the basic health insurance market in the Netherlands, would gladly accept 0.1 as a ratio, whereas more profit-oriented markets, as in the United States, may strive for greater ratios (e.g., 0.5).

By converting Equation 3.1 (section 'Variation in uncertainty of financial return') to Equation 3.2, the desired profit mark-up can be found for any number of individuals from a risk group. In this analysis, the return on a risk-free asset, *Rf*, is set to €0 to simplify comparison. We calculate the mean per-person profit mark-up (PMU), for *N* consumers contracted from risk group *g*, as:

$$PMU_{g,N} = S * Standard\ Deviation\left(\bar{r}_g,\ N\right) \qquad (3.2)$$

with S the selected Sharpe ratio and the standard deviation of mean residual spending in risk group *g*, and the number of individuals, *N*, selected from that group *g*. The standard deviation of the mean residual spending in risk group *g* for a given *N* ($Standard\ Deviation\left(\bar{r}_g,\ N\right)$) is found through Equation 3.3:

$$Standard\ Deviation\left(\bar{r}_g,\ N\right) = \frac{Standard\ Deviation(r_g)}{\sqrt{N}} \qquad (3.3)$$

where the standard deviation of the risk group of individuals (or: individual-level standard deviation) in group *g* ($Standard\ Deviation(r_g)$) is divided by the square root of the number of individuals, *N* (i.e., the portfolio size). By doing so, the degree of variation for the number of individuals of the risk group is set to reflect the impact of the law of large numbers, which decreases variation. For the analyses it is important to understand the difference between the two metrics of standard deviation.

Combining Equations 3.2 and 3.3 facilitates the calculation of the mean per person profit mark-up for individuals of any risk group. The differences in mean per person profit mark-up across groups indicates the selection incentives under a community-rated premium.

Next, we derive the safety mark-up for group *g*, i.e., the mark-up that insurers would have to charge to fulfill the capital requirements. Assume an insurer j, places an insurance plan on the market and, as a consequence, attracts a certain number of individuals from group *g*. EU Solvency II legislation (footnote 4) specifies what capital requirements (*SCRj*) result from that particular chain of events through Equations 3.4 and 3.5 as follows:

$$SCR_j = 3 * \delta_j * V_{total,\ j} \tag{3.4}$$

Here, *Vtotal, j* is a summed measure[8] for the volume of reserves of the insurer ($V_{reserves,\ j}$) and the annual income through premiums and risk equalization payments ($V_{incomes,\ j}$), the latter of which is directly influenced by the composition of the portfolio.[9] Moreover, δ_j reflects a weighted parameter, based on weights set by the EU and the volume measures that make up $V_{total,\ j}$. Equation 3.5 defines the computation of δ_j as follows:

$$\delta_j = \frac{\sqrt{\delta^2_{incomes} * V^2_{incomes,j} + \delta_{incomes} * V_{incomes,j} * V_{reserves,j} + \delta^2_{reserves} * V^2_{reserves,j}}}{V_{incomes,\ j} + V_{reserves,\ j}} \tag{3.5}$$

Here, *δincomes* is set to 2.7%, a defined weight by the European Union, specific to the Dutch health insurance market (EUR-Lex, 2015b).[10] $\delta_{reserves}$ is legally set to 5% (footnotes 4 and 10). Note that Equations 3.4 and 3.5 can be integrated but are rather presented as

8 $V_{total,\ j} = V_{incomes,\ j} + V_{reserves,\ j}$. Moreover, in our estimations we assumed zero financial reserves for insurer j, setting $V_{reserves,\ j}$ to €0. These reserves should be amassed eventually. Moreover, Table A3.1. in the Appendix provides an overview of the effect of various levels of financial reserves on our estimations.

9 As a consequence of this legal structure, the potential enrollment of L versus H could generate additional risk selection incentives in the extraordinary situation where both the mean residual spending and its standard deviation are exactly equal between the two groups with different mean risk equalization payments. Since the SCR is directly influenced by these predicted costs, contracting an individual from the risk group with relatively high mean risk equalization payments results in a larger increase to the amount of capital legally required to be retained, ceteris paribus. In theory, the absence of deviations in predicted residual spending, and the variation thereof, between groups may still be insufficient to abolish risk selection incentives for insurers.

10 Article 1 of the Commission Implementing Regulation (EU) 2015/2013 of 11 November 2015 states that due to the health risk equalization system in place in the Netherlands, $\delta_{incomes}$ is set to 2.7% rather than 5% in other Member States. Since risk equalization payments decrease the degree of risk borne by insurers, the SCR_j is adjusted accordingly.

in the official legislation.[11] Although there is no parameter for portfolio size, this is taken into account through the total volume of income.

Through Equations 3.4 and 3.5 we can now find the legally required amount of capital (*SCRj,g,N*) that an insurer would need to hold when *N* individuals from group *g* are contracted. Moreover, by dividing that capital requirement by *N* we find the mean per person capital requirements (*CR*) for group *g* as follows:

$$CR_g = \left(\frac{SCR_{j,g,N}}{N}\right) \tag{3.6}$$

Last, to convert the per-person amount of capital required for group *g* to a safety mark-up (SMU) for group *g*, the CR_g should be multiplied by parameter ρ for the 'cost of capital' (e.g., opportunity costs or interest rate on loans) (Equation 3.7). For example, a higher safety mark-up might require the insurer to either maintain a larger sum of own capital resources, effectively reducing the amount of other types of capital, or to attract new capital, which results in additional loans. It should be noted that Solvency II for SCR requires that most of the capital to be held by insurers should be own capital, restricting the options for loans or other types of assets. In our analyses, we apply a 5% and 10% rate for ρ to reflect the cost of capital as an indication.[12] The difference in the resulting safety mark-ups between risk types provides an approximation of selection incentives in community-rated markets, presented by the earlier example of a 10% interest rate on the €400 capital requirements.[13]

$$SMU_g = \rho * CR_g \tag{3.7}$$

3.3.4 Quantifying the effects of risk-sharing on heteroscedasticity of residual spending

In addition to risk equalization, many health insurance markets also include some form of risk-sharing (which provides insurers with additional payments based on the actual spending of insured) (Brammli-Greenberg et al., 2019). One common form of risk-sharing is outlier risk-sharing (Van Barneveld et al., 2001b), as applied in Germany and the US Marketplaces system (Layton et al., 2018c; McGuire et al., 2021a). Outlier risk-sharing means that insurers are compensated ex-post by the regulator for (a proportion of) individual-level spending above a predefined threshold. An important motive for such risk-sharing is to protect insurers from the risk of large losses (McGuire & Van Kleef, 2018b). Outlier risk-sharing is likely to mitigate the heteroscedasticity in residual spending. Therefore,

11 The main point of the notation is that the variation in *SCR* is mainly determined by $V_{total,\ j}$ and to a smaller extent by δ_j.

12 The European Insurance and Occupational Pensions Authority (EIOPA) applies a cost of capital rate of 6% for the Solvency II legislation (EIOPA, 2015).

13 Alternatively, an insurer could build up the required capital by raising overall premiums to enrollees. However, this route is disregarded in this exercise.

as a final step in our analysis, we examine the effect of risk-sharing on our measures of interest. More specifically, we simulate the effects of a common form of outlier risk-sharing: 80% cost-based compensation of individual-level spending above a threshold of €100,000.[14] In line with international standards, the presence of outlier risk-sharing is taken into account in the estimation of the risk equalization model. This means the model is recalibrated on spending *net of* outlier risk-sharing, thereby reducing the total amount of risk equalization payments. As a result, the ex-post compensations are budget-neutral: no extra funds flow into the system. These compensations are, therefore, regarded as incomes to insurers, just as payments from the risk equalization system would be.

3.4 RESULTS

This section presents the results of our analyses. We first report some descriptive statistics of the dataset available for this study (section 'Descriptive statistics'). Next, we demonstrate the measures of variance that serve to quantify the heteroscedasticity of residual spending (section 'Quantifying the heteroscedasticity of residual spending'). In the section 'Quantifying selection incentives by simulating group-level profit and safety mark-ups', the two different mark-ups are calculated for a variety of risk groups and in the final part (section 'Effects of risk-sharing on heteroscedasticity of residual spending'), the effect of risk-sharing on our measures of interest is demonstrated.

3.4.1 Descriptive statistics

Table 3.1 presents an overview of the various risk groups of individuals chosen for our analyses, representing different levels of prevalence in the population and diverging levels of mean spending. In addition to groups based on age and morbidity adjuster qualification, three other separations have been made. First, diabetics reflect a relatively large group of chronically ill individuals. The cluster of conditions grouped in DCG10 covers a fairly sizeable group (roughly 67,000 individuals) with a substantial level of average spending. Last, the final group of the PCG adjuster is shown to reflect the costliest type of individuals recognized by the risk equalization model: those that use specific and extremely costly pharmaceuticals. For each of these groups, mean *residual* spending equals €0 (as a consequence of the ordinary least squares method used to estimate the payment weights of the risk equalization model).[15] This implies that – despite substantial differences in mean spending across risk groups – there are no predictable profits/losses for these groups, implying the absence of risk selection incentives based on differences in mean residual spending between these groups.

14 This form of outlier risk sharing is actually applied in the German sickness fund insurance (McGuire et al., 2021a).

15 Since individuals can quantify for multiple DCGs and more than once for the same class, the mean residual spending for DCGs is different from €0. Moreover, all PCG risk-classes have a mean of €-1.

Table 3.1. Descriptive statistics of risk groups

Risk group	Prevalence (%)	Mean actual spending 2018 (€)	Mean predicted spending 2018 by risk equalization (€)	Mean residual spending 2018 net of risk equalization (€)
Entire population	100	2,408	2,408	0
Individuals aged 0-64	80.7	1,600	1,600	0
Individuals aged 65+	19.3	5,792	5,792	0
No morbidity flags	74.8	1,061	1,061	0
≥ 1 morbidity flags	25.2	6,409	6,409	0
Diabetics (PCG12-15)	3.4	8,407	8,408	-1
Cluster of conditions (DCG10)	0.4	28,532	28,495	37
Extremely costly pharmaceutical usage (PCG38)	0.0002	602,637	602,638	-1

Note: N = 17,256,295, corresponding to 16,951,700 insured years. Frequencies are calculated as a percentage of total insured years in the population. The risk groups for morbidity flags are derived from grouping all insured that do not qualify for compensation through any of the six morbidity adjusters (i.e., 'healthy individuals') and by taking the complementary group (i.e., 'unhealthy individuals').

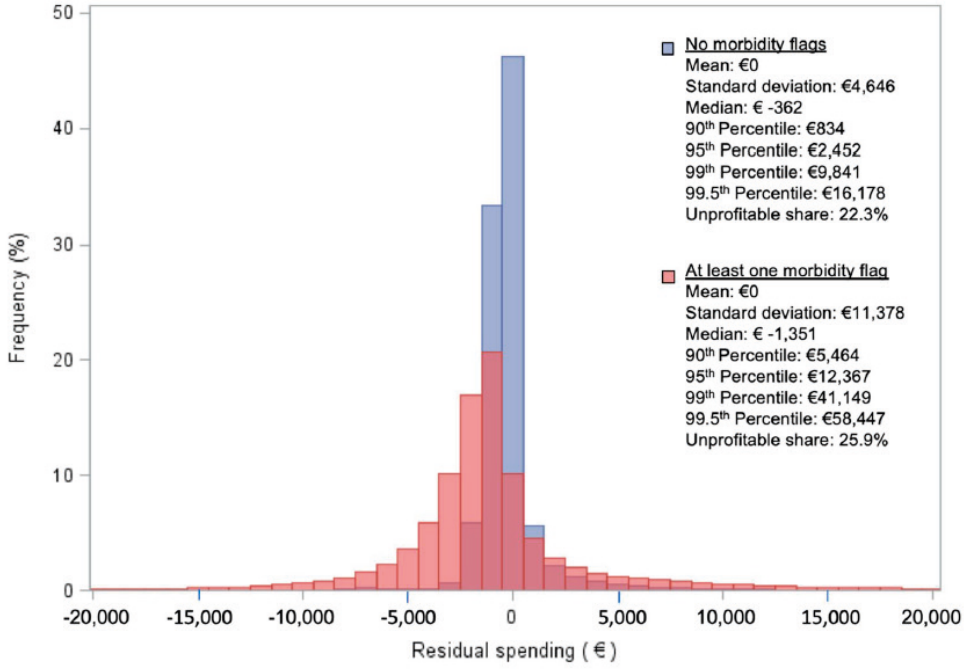
3.4.2 Quantifying the heteroscedasticity of residual spending

To get a first glimpse of the heteroscedasticity of residual spending, Figure 3.1 presents the distribution of residual spending for two risk groups as follows: 1) individuals without a morbidity flag (blue) and 2) individuals with at least one morbidity flag (red). Although the graph ranges from €-20,000 to €20,000, the residual spending in the extremes reaches well beyond those borders. Nevertheless, roughly 99% of all residuals are captured within the presented range. Note that positive residual spending indicates a loss to insurers, whereas a negative residual represents a profit.

Although the mean residual spending equals €0 for both risk groups, the distributions are quite distinct. In addition to the visual contrast between the two distributions, some metrics are presented to summarize the divergence. The red distribution has a standard deviation of €11,378, far larger than that of the total population (€6,978), while the blue distribution has a smaller standard deviation of €4,646. The results for the group without morbidity are more concentrated around €0 than those for the group with morbidity, implying that insurers face more uncertainty regarding the ex-post financial result for individuals with a morbidity flag than for those without, despite the fact that the mean residual spending equals zero for both groups. Last, the relative share of individuals that lead to a loss for insurers is greater for the group with a morbidity flag, implying a greater chance of a financial loss when contracting an individual from this group.

Figure 3.2 dives deeper into the group with at least one morbidity flag and presents the distributions of residual spending for groups defined by the number of morbidity flags.

Figure 3.1. Distribution of individual-level residual spending, separately for individuals without any morbidity flag (*N* = 12,682,389) and those with at least one morbidity flag (*N* = 4,269,311)

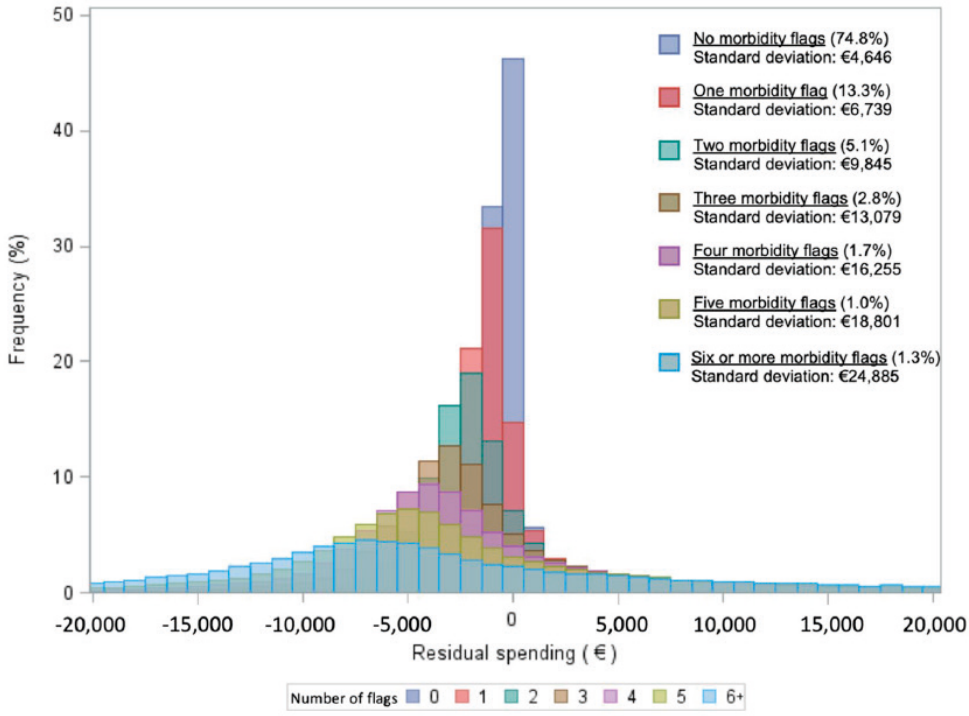


While the 2021 risk equalization model has six types of morbidity adjusters, consumers can qualify for multiple flags within some of these adjusters, meaning that individuals can end up with more than six morbidity flags. The clear takeaway from Figure 3.2 is that the variation of residual spending increases with the number of morbidity flags.

While Figures 3.1 and 3.2 present the distributions of residual spending at the level of individual enrollees, Figure 3.3 presents the distribution of *mean* residual spending at the level of fictive portfolios, resulting from the simulations. Four portfolio sizes are considered: 1,000, 10,000, 100,000, and 1,000,000 enrollees. For every combination of 'no morbidity flag' versus 'at least one morbidity flag' and the four portfolio sizes, a thousand random draws with replacement have been done to approximate the distribution of *mean* residual spending. In addition to the visual presentation, the variation and 99.5th percentile of mean residual spending decrease with portfolio size, as expected. Taking the portfolio size of 10,000 as an example, the 99.5th percentile is €152 higher for the red group (€264) than for the blue group (€112), while the standard deviation (€104) is €63 higher for the red group than for the blue one (€41). These values demonstrate the heteroscedasticity of residual spending that remains when the portfolio size is accounted for.

To supplement the results shown in Figure 3.3, Table 3.2 presents the outcomes for the other groups from Table 3.1. Although all risk groups have a mean residual spending of €0

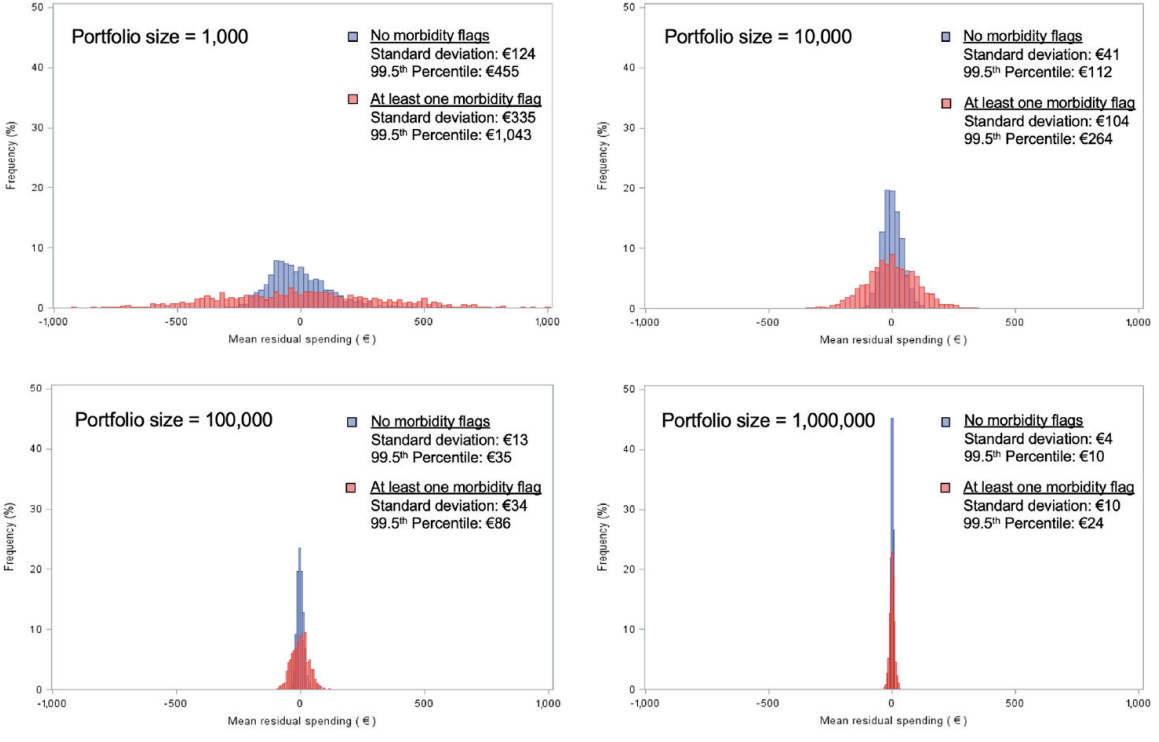
Figure 3.2. Distribution of individual-level residual spending, separately for groups of individuals based on the number of morbidity flags in the risk equalization model.



Note: The mean residual spending of these groups is not exactly €0 (ranging between €- 169 and €49), which is the result of the fact that the number of morbidity flags is not an explicit risk adjuster in the model. Therefore, the average result is not by definition set to €0. Through a small fixed correction (that does not affect the variation in residual spending), the means are set to €0 to concentrate the distributions around the same value for ease of presentation. For instance, if the average result for the group of individuals with precisely two morbidity flags is €49, we lowered the result for all individuals of that particular risk group by a fixed amount of €49.

(considering footnote 15), their distributions of residual spending vary. Both the standard deviation and the 99.5th percentile are notably different for groups with relatively low spending compared to those with relatively high spending. Moreover, the two metrics are derived for portfolio sizes of 10,000 enrollees, similar to the simulations in Figure 3.3. The outcomes indicate that insurers face greater uncertainty in the mean residual spending for individuals qualifying for DCG10 (€287) than for 'healthy' individuals (€41). The 99.5th percentile of residual spending for the DCG10 group is also considerably higher than for the group without any morbidity flag (€741 versus €126), indicating that insurers would eventually need more financial reserves to endure the 1/200 risk for a portfolio consisting

Figure 3.3. Distribution of mean residual spending for 1,000 simulations of portfolio sizes of 1,000, 10,000, 100,000, and 1,000,000 individuals, separately for people with at least one morbidity flag and those without a morbidity flag.



Note: For each combination of portfolio size and risk group, a thousand random draws with replacement have been performed. The figures show the distribution of mean residual spending for these 1,000 draws.

Table 3.2. Measure of variance in residual spending for eight risk groups

			Portfolio size (*N*) = 1		Portfolio size (*N*) = 10,000	
Risk group	Total insured years	Mean actual spending (€)	Standard deviation (€)	99.5th percentile (€)	Standard deviation (€)	99.5th percentile (€)
Entire population	16,951,600	2,408	6,798	33,352	65	191
Individuals aged 0-64	13,685,072	1,600	5,888	21,371	53	151
Individuals aged 65+	3,266,628	5,792	10,352	55,371	97	247
No morbidity flags	12,682,389	1,061	4,646	16,178	41	126
≥ 1 morbidity flags	4,269,311	6,409	11,378	58,447	106	271
Diabetics (PCG12-15)	538,810	8,407	12,234	64,470	123	300
Cluster of conditions (DCG10)	67,345	28,532	29,957	152,530	287	741
Extremely costly pharmaceutical usage (PCG38)	27	602,637	350,078	819,092	*n/a*	*n/a*

Note: For the portfolio size of 10,000, a thousand simulations of random draws with replacement have been performed for the relevant risk group. The standard deviation and 99.5th percentile of residual spending presented are derived from those one thousand means. The risk groups for morbidity flags are derived from grouping all insured that do not qualify for compensation through any of the six morbidity adjusters ('healthy individuals') and by taking the complementary group ('unhealthy individuals'). The portfolio size of 1 serves to demonstrate the uncertainty/risk when randomly attracting one individual from a specific risk group. The portfolio size of 10,000 serves to demonstrate the uncertainty/risk regarding the mean per-person financial result when randomly attracting 10,000 individuals from a specific risk group.

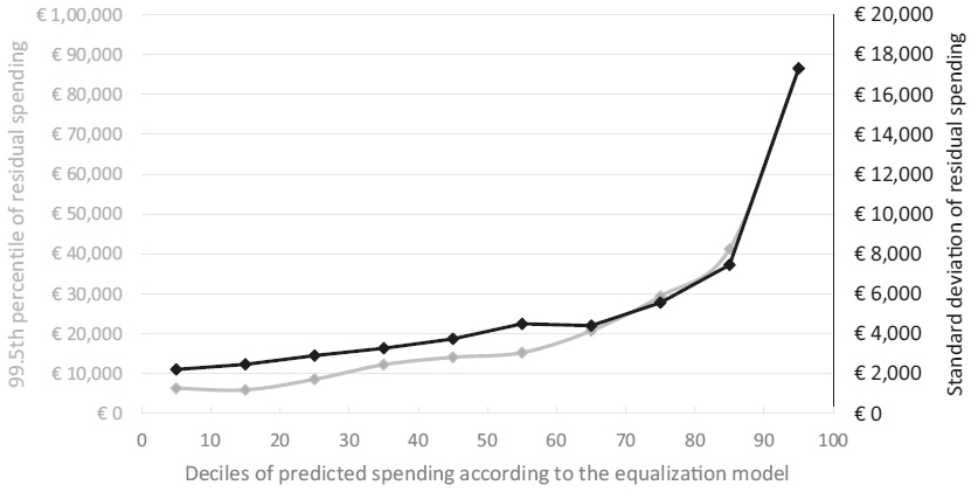
of 10,000 people from DCG10 compared to a portfolio consisting of 10,000 people from the 'healthy' group.

For PCG38 it was not meaningful to simulate a portfolio size of 10,000 since – in the entire Dutch population – this group consists of no more than 27 individuals. Nevertheless, the existence of such small but very expensive risk groups is highly interesting in the light of the selection problems analyzed in this paper. For bigger groups, insurers have better opportunities to exploit the law of the large numbers by enrolling more individuals. For a group like PCG38 these opportunities are absent: even when insurers would enroll the entire group of people with PCG38, the standard deviation and 99.5th percentile of mean residual spending within the group remain enormous.

The results from Table 3.2 demonstrate an apparent positive correlation between the mean spending of risk groups and the two applied metrics of variation of *residual* spending. To further illustrate this correlation, Figure 3.4 shows the standard deviation and 99.5th percentile of residual spending for deciles of individual-level predicted spending according to the risk equalization model. Individuals with low predicted spending are on the left, and those with the highest predicted spending are on the right. For the standard deviation and the 99.5th percentile of residual spending, a similar pattern is found by using

two separate scales on the right and left y-axes, respectively. The clear, non-linear trend observed in the graph demonstrates the substantial concentration of uncertainty in residual spending and risk of large losses among those with the highest predicted spending.

Figure 3.4. Measures of variance in residual spending (standard deviation in black; 99.5th percentile in light gray) per decile of the population ranked by predicted spending according to the risk equalization model



Note: All insured were ranked by their level of predicted spending according to the risk equalization model. The scales used for the two y-axes are unequal to facilitate comparison in the trends.

3.4.3 Quantifying selection incentives by simulating group-level profit and safety mark-ups

To quantify the potential selection incentives that result from the observed heteroscedasticity of residual spending, this section demonstrates the theoretical mark-ups that insurers would have charged to individuals of the different risk groups in a risk-rated market. The motivation for this exercise is that differences in mark-ups among groups provide an indication of selection incentives regarding these groups in a social health insurance market that typically includes community-rating, an acceptance policy for insurers, and a defined benefit package. First, the uncertainty of residual spending (standard deviation) is converted to a profit markup based on the Sharpe ratio, and second, the safety markup is calculated based on EU solvency legislation and assumptions about the cost of capital.

Table 3.3 presents the results of the conversion of the standard deviation of the eight respective risk groups to a per-contract profit mark-up, derived from specific values of

Table 3.3. Uncertainty of residual spending, the profit mark-up in a risk-rated market, and the corresponding selection incentives in a community-rated market

			Portfolio size (N) = 10,000			
			Sharpe ratio = 0.1		Sharpe ratio = 0.5	
Risk group	Standard deviation, σ (€)	$\frac{\sigma}{\sqrt{10{,}000}}$ (€)	Profit mark-up in risk-rated market (€)	Selection incentives in community-rated market (€)	Profit mark-up in risk-rated market (€)	Selection incentives in community-rated market (€)
Entire population	6,798	70	7	–	35	–
Individuals aged 0-64	5,888	59	6	+ 1	29	+ 6
Individuals aged 65+	10,352	104	10	– 3	52	– 17
No morbidity flags	4,646	46	5	+ 2	23	+ 12
≥ 1 morbidity flags	11,378	114	11	– 4	57	– 22
Diabetics (PCG12-15)	12,234	122	12	– 5	61	– 26
Cluster of conditions (DCG10)	29,957	300	30	– 23	150	– 115
Extremely costly pharmaceutical usage (PCG38)	350,078	3,501*	350*	– 343	1,750	– 1,715

Note: The PCG38 group only contains 27 insured years; the asterisks (*) for this group indicate that not 10,000 but 27 insured years are contracted. The selection incentives for the risk groups are derived by subtracting the profit mark-up for the respective risk group from the mark-up for the average individual in the population. This simulates a community-rated premium for the overall population and the expected selection incentives for the respective risk groups.

the Sharpe ratio and the number of contracted individuals. The third column shows a hypothetical portfolio of 10,000 individuals from the specific risk groups (except for PCG38, where the risk group includes no more than 27 individuals, shown by the asterisk in Table 3.3). In the final four columns of the table, two different values of the Sharpe ratio are used for the various portfolios to simulate either a profit mark-up for insurers on a risk-rated market or the selection incentives for insurers on a community-rated market. First, the standard deviation is divided by the square root of the number of enrollees to facilitate the simple application of the Sharpe ratios (Equation 3.3). By multiplying the result with a specific Sharpe ratio, the per-contract profit mark-up for all of the 10,000 contracts from any risk group can be derived. Moreover, by setting the resulting per capita mark-up for the average individual as a benchmark, an approximation of the selection incentives for insurers towards the individuals from the different risk groups can be simulated.

Thus, to cover for the difference in uncertainty of financial return between young and elderly individuals, a risk-rating insurer could seek to charge the latter group €10 per contract while requiring €6 per contract from the young, complementary group (assuming a conservative Sharpe ratio of 0.1 for not-for-profit markets). However, taking a less conservative Sharpe ratio of 0.5 and a more uncertain group as DCG10 results in a substantial difference of €127 between that specific group (€150) and the group without a morbidity

flag (€23). Notably, the mark-up for the PCG38 group would be substantial, underscoring the severe underlying financial risk for insurers regarding this risk group. Last, by taking the difference between the mark-up for an average individual in the total population (€7 or €35 for a Sharpe ratio of 0.1 or 0.5, respectively) and that for the average individual in risk group *g*, the incentives for risk selection regarding risk group *g* can be simulated for insurers on a community-rated market.

Table 3.4 provides an indication of how the risk of ruin translates to a safety mark-up. The second column presents the mean predicted spending of the risk groups, which reflects the mean per-contract income to insurers. This income – which is composed of the annual premium paid by enrollees plus the risk equalization payment received from the regulator – is crucial in determining the safety mark-up (see section 'Quantifying the selection incentives generated by heteroscedasticity of residual spending' for further detail). The third column shows the per-contract contribution to the solvency capital requirements for an insurer, subject to the assumptions discussed in the Methods section. For instance, a portfolio of diabetes patients results in a capital requirement of €681 per person, while a portfolio of individuals without any morbidity flag comes with a capital requirement of €86 per person. Moreover, comparing the age groups results in a difference of €339 between the young (€130) and elderly (€469). Using two values for the parameter for the 'cost of capital', ρ, we find the respective safety markups. Assuming a cost of capital of 10% relative to the capital requirement, the safety markup for an individual in risk group DCG10 is €211 greater than for an average individual. The differences between most other groups are smaller but non-negligible. Similar to the findings in Table 3.3, the difference between the respective mark-up for an average insured individual and that for one from a specific risk group provides an indication of the selection incentives for insurers on a community-rated market.

Although the SCR does not consider any metric of variation, the level of predicted spending is used, which is positively correlated with the variation of residual spending, as previously shown by Figure 3.4. Moreover, as discussed before, the SCR hardly considers the portfolio size. Whether the portfolio size equals 1,000, 10,000, or 100,000 individuals, the safety mark-ups per contract for those individuals remain equal. The differences between risk groups are therefore retained. These results clash with those presented in Figure 3.3, demonstrating that the risk substantially decreases with group size through the law of the large numbers.

3.4.4 Quantifying the effects of risk-sharing on heteroscedasticity of residual spending

As a final step in our analysis, a risk-sharing modality typical for health insurance systems is simulated to concisely demonstrate the impact of such policies on the heteroscedasticity of residual spending. Table 3.5 demonstrates the effect of outlier- risk-sharing – with

80% compensation of individual-level medical spending above a threshold of €100,000 – on the two metrics of heteroscedasticity and the corresponding mark-ups. With most of the costs cut off above the threshold and the risk equalization model recalibrated, the variation in residual spending decreases. Consequently, the profit markup decreases too.

While the 99.5th percentile of residual spending, the risk of ruin, reduces as a consequence of risk-sharing, the safety mark-up is unaffected. The fact that the budget is kept neutral in our analysis results in a similar income for the insurers with an average risk profile. Although the overall risk (variation) is reduced through the risk-sharing efforts, the SCR does not directly incorporate that effect. The safety mark-up remains equal because the solvency regulations for capital requirements (see Equation 3.5) only include the mean spending of an insurer population and not the variation in residual spending. Since the

Table 3.4. Risk of ruin and the theoretical safety mark-up for eight risk groups

			$\rho = 5\%$		$\rho = 10\%$	
Risk group	Mean predicted spending (€)	Per i contribution to SCR_j (€)	Safety mark-up in risk-rated market (€)	Selection incentives in community-rated market (€)	Safety mark-up in risk-rated market (€)	Selection incentives in community-rated market (€)
Entire population	2,408	195	10	–	20	–
Individuals aged 0-64	1,600	130	7	+ 3	13	+ 7
Individuals aged 65+	5,792	469	23	– 13	47	– 27
No morbidity flags	1,061	86	4	+ 6	9	+ 11
≥ 1 morbidity flags	6,409	519	26	– 16	52	– 32
Diabetics (PCG12-15)	8,408	681	34	– 24	68	– 48
Cluster of conditions (DCG10)	28,495	2,311	116	– 106	231	– 211
Extremely costly pharmaceutical usage (PCG38)	602,638	48,814	2,441	– 2,431	4,881	– 4,861

Note: The safety mark-up is derived through Equations 3.4–3.7 from the Methods section. The selection incentives for the risk groups are derived by subtracting the safety mark-up for the respective risk group from the mark-up for the average individual in the population. This simulates a community-rated premium for the overall population and the expected selection incentives for the respective risk groups.

Table 3.5. Measure of variance in mean residual spending and the profit and safety mark-ups for two risk groups in a simulation with and without risk-sharing supplementing risk equalization

	Std deviation (*N*) = 10,000 (€)		PMU (*N*) = 10,000 (€)		99.5th percentile (*N*) = 10,000 (€)		SMU (*N*) = 10,000 (€)	
	Risk-sharing		Risk-sharing		Risk-sharing		Risk-sharing	
Risk group	No	Yes	No	Yes	No	Yes	No	Yes
No morbidity flags	41	35	5	4	126	92	9	9
≥ 1 morbidity flags	106	96	11	9	271	217	52	52

Note: For the portfolio size of 10,000, a thousand simulations of random draws with replacement have been performed for the relevant risk group. The standard deviation and 99.5th percentile of residual spending presented are derived from those 1,000 means. The profit and safety mark-up are calculated as discussed in the Methods section, so the profit mark-up is based on a different standard deviation than shown in the first column. A Sharpe ratio of 0.1 is used in the calculation of the profit mark-up and a cost of capital of 10% for the safety mark-up. The risk-sharing modality implies that 80% of costs of enrollees above €100,000 are compensated, and the risk equalization model is recalibrated to this adjustment. The values in the final columns are equal to demonstrate how the current implementation of the SCR would not incorporate the reduction of risk through risk-sharing and that the level of alteration to the inherent parameters that could be considered to do so is unknown.

risk equalization payments are counted as incomes, the SCR increases accordingly, leading to more capital to be held by insurers. However, if the regulator were to implement a risk-sharing modality as a supplement to risk equalization, the parameters in Equations 3.4 and 3.5 can be altered by the governing body to reflect the decreased degree of risk borne by insurers. Nevertheless, in line with the expectations, the risk-sharing modality commonly applied in health insurance markets reduces heteroscedasticity of residual spending.

3.5 SIZE OF THE PROBLEM AND POTENTIAL SOLUTIONS

In this section we discuss the extent to which the selection incentives caused by heteroscedasticity of residual spending are problematic. The extent of the problem depends on two factors: 1) the size of the incentives and 2) the options for insurers to engage in risk selection. Additionally, we explain various strategies that could be considered to mitigate the selection incentives due to heteroscedasticity of residual spending.

3.5.1 The extent of the problem of risk selection incentives

Through our simulations, we illustrated how heteroscedasticity of residual spending after risk equalization confronts insurers with selection incentives. Using Dutch health insurance data of the total population, we find considerable differences in the standard deviations of residual spending between different risk groups of insured despite the fact that the mean residual spending equals €0 for the considered groups. We simulated how – in a risk-rated market – this heteroscedasticity would most likely have resulted in differ-

ent profit and safety markups. For example, for two mutually exclusive groups based on morbidity status – i.e., individuals without any morbidity flag in the equalization model versus those with at least one morbidity flag – we found a difference in profit mark-ups of €6 for a conservative Sharpe ratio of 0.1 and €34 for a less-conservative Sharpe ratio of 0.5. These findings are, however, subject to assumptions on portfolio size: smaller sample sizes increase uncertainty and thereby mark-ups. Insurance markets with larger portfolios will therefore endure less concerns of uncertainty. Nevertheless, for specific risk groups of individuals with few overall members (such as PCG38 in our analyses), uncertainty will always be large. Moreover, our simulation of the EU solvency legislation indicates a difference in safety mark-ups between the groups with and without at least one morbidity flag of €22 with the 'cost of capital' parameter set to 5% and €43 when that parameter is doubled to 10%. Taking the profit and safety mark-ups together creates a per capita difference of €77 (€34 profit mark-up and €43 safety mark-up) between the groups with and without at least one morbidity flag. These findings imply that on a community-rated insurance market (where insurers must charge the same premium to these groups), contracting an individual from the non-morbidity group is €77 more appealing to insurers than contracting one from the complementary group of individuals with at least one morbidity flag. Clearly, greater differences can be found when comparing 'healthy' individuals with specific risk groups for which the variation in residual spending is (much) larger than for the total group with a morbidity flag. The size of these selection incentives, however, strongly depends on the parameters and assumptions underlying our analyses. In general, the selection incentives (as quantified above) will be greater (smaller) for markets with higher (lower) demands for insurance firm profit-return or stricter (looser) solvability requirements by regulators. As for the Dutch case, however, the simulated selection incentives that result from heteroscedasticity of residual spending can be considered nontrivial.

However, the fundamental question remains: to what extent do the incentives evolve into actions and thus become problematic? Although direct rejection of prospective enrollees is prohibited in many health insurance systems (due to open enrollment requirements), a variety of actions of risk selection can be legally undertaken to distort the natural enrollment of individuals (Douven & Van de Ven, 2022b; Van Kleef et al., 2022a). One example of risk selection is insurers not contracting the best care for enrollees with a (specific) chronic disease that is known to be unprofitable. Other examples include offering high premium discounts for the uptake of a voluntary deductible, the design of supplementary insurance plans, and selective marketing towards groups that are known to be profitable (Van de Ven et al., 2015).

Selection activities are often subtle, and providing clear evidence of risk selection is not straightforward (Van de Ven et al., 2017). One way to demonstrate risk selection is to compare insurer-level residual spending net of risk equalization payments for two

consecutive years. The glaring problem, however, is that any differences might also be related to other factors, such as efficiency. Nevertheless, *signals* of risk selection, as in the forms discussed above, are prevalent and provide an indication of insurers exploiting the knowledge on predictable profitability of risk groups. For example, such signals include the offering of so-called 'twin products' and the targeting of specific groups (such as highly educated people).[16] Other signals include complaints by insurers that the predictable losses on chronically ill people discourage them from organizing the best care for chronically ill people (Van Kleef et al., 2022a).

3.5.2 Solutions to the problems that result from heteroscedasticity of residual spending

If the heteroscedasticity found in this study is considered problematic for the functioning of the insurance market, the current focus of risk equalization design (i.e., compensating for differences in mean spending across risk types) is insufficient. Even if the risk equalization model would perfectly compensate insurers for differences in mean spending across risk types, heteroscedasticity in residual spending remains. In general, we see three potential solutions to mitigate the selection incentives that result from heteroscedasticity in residual spending.

A first solution may be to supplement risk equalization with risk-sharing. In this study, we simulated how a common form of risk-sharing, outlier risk-sharing, reduces heteroscedasticity of residual spending. These findings imply that risk-sharing indeed provides the regulator with an instrument to mitigate the type of selection incentives studied in this paper. However, risk-sharing also comes with a price: incentives for insurers to control costs are reduced. Recent studies on innovative forms of risk-sharing demonstrate how the tradeoff between selection and efficiency can be mitigated: targeting the ex-post compensations to individuals with high residual spending instead of high spending (McGuire et al., 2020; McGuire et al., 2021a). Moreover, ex-post compensations for high residual spending can be financed by ex-post repayments for extremely low residual spending. Targeting the extremes in residual spending cuts off the wide edges of the distribution of residual spending, reducing heteroscedasticity. Alternatively, risk-sharing could be applied conditionally on specific risk groups, e.g., those groups with the highest variance in residual spending.

The second solution could be to modify the risk equalization payments and compensate insurers ex-ante for the heteroscedasticity in residual spending. For example, the regulator could estimate the selection incentives due to heteroscedasticity in residual spending (following the methods of this paper) and modify the risk equalization payments for risk

16 Twin products are nearly-identical basic-insurance products with different prices in combination with different options for supplementary products. In general, the lower-priced twin comes with less generous (supplementary) coverage than the higher-priced twin

groups to eliminate these incentives. In general terms, this means more (less) compensation for risk groups with a relatively high (low) variance in residual spending. Returning to the findings of our simulations, the difference in mark-ups between risk groups could be used as a basis for modifying the risk equalization payments. Taking the results in Table 3.4 as an example, the equalization payments for individuals without a morbidity flag could be decreased by €11, while the payments for the complementary group with morbidity flags could be increased by €32. While this strategy would render risk groups that are potentially vulnerable to actions of risk selection more financially appealing to insurers, the design – like our research – involves assumptions with respect to Sharpe ratio and financial reserves of insurers. To some extent these assumptions will be surrounded with uncertainty, which creates the potential of under/overshooting. Further research is required to determine the optimal compensations to adequately alleviate the risk selection incentives from heteroscedasticity.

A third solution could be to permit limited differentiation in mark-ups in premium-setting across individuals, diminishing the risk selection incentives at the cost of health insurance affordability and solidarity. With perfect equalization – as assumed in this study – the mark-up will reveal the different profit and safety mark-ups that insurers require for different risk types. The advantage of this approach over the second solution (as discussed in the previous paragraph) is that premium differentiation does not require any (arbitrary) assumptions to be made by the regulator about hypothetical profit and safety mark-ups. However, existing risk equalization models do not yet perfectly compensate for differences in mean spending across risk types. Allowing risk-rated premiums could lead insurers to not only reflect differences in profit and safety mark-ups across groups but also differences in predictable profits and losses across these groups. This could exacerbate the negative impact on affordability and fairness.

All in all, we expect the uncertainty of the financial result on a group level to result in selection incentives for competing insurers, even when the risk groups are correctly compensated for their mean expenditure. The actions that insurers in regulated markets may undertake to evade the enrollment of individuals from 'risky' groups are problematic to the functioning of the insurance system. In the hypothetical scenario with perfect risk equalization on a group level, various solutions are possible to decrease the heteroscedasticity in residual spending and the resulting selection incentives, but all come with respective challenges and tradeoffs. In the next, and final, section we reflect on the implications of our research as a whole and to what extent intervention is required.

3.6 DISCUSSION

Our conclusion that residual spending in Dutch basic health insurance is subject to substantial heteroscedasticity follows directly from the data. However, the subsequent quantification of selection incentives in a community-rated market (through simulation of the profit and safety mark-up in a free market) incorporates assumptions. The profit mark-up is built upon the Sharpe ratio, a metric used in investment risk management and not specifically designed for health insurance markets with mostly not-for-profit insurers, as in the Netherlands. Despite the use of a conservative value for the Sharpe ratio in our study, it should be further explored how this metric relates to the actual risk health insurers face in practice. In for-profit markets the Sharpe ratio of competing insurers, or any other approximation of profit-seeking behavior, is likely to be higher than in nonprofit markets. More generally, the strength of the incentives resulting from heteroscedasticity does not only depend on the size of the loadings that we identify but also on the behavior of the insurers and how they respond to financial incentives. Obviously, the interpretation of the results is subject to such uncertainty.

The calculations of the safety mark-up depend upon a baseline level of financial reserves (i.e., the financial reserves insurers already possess before the start of a new contract period), or rather the absence thereof. Additional analyses show that the magnitude of the safety mark-up, based on the solvency capital requirements, remains relatively similar when using different starting levels of financial reserves (Table A3.1 in the Appendix demonstrates the effect of different levels of financial reserves on the resulting safety mark-ups for our considered risk groups). Remarkably, an *increase* of the baseline level of financial reserves results in an *increase* of the mark-up (or the contribution to the capital requirement), and the increase of the mark-up is nearly equal between the various risk groups analyzed, despite the differences in 'risk' (average expenditure of the group). The overall increase is more or less linear, resulting from the increase to financial reserves alone and almost entirely irrespective of the 'risk' of the respective risk groups. In practice, however, the initial level of financial reserves may play a more crucial role. A substantial share of the mandatory capital for the SCR is required to be an insurer's own, while in our analysis we used only the 'cost of capital' to quantify the selection incentives. However, insurers with insufficient financial reserves may have to charge an additional fee to raise their own resources in addition to the costs of loans or alternative opportunity costs. As a result, insurers with different levels of financial reserves might have to charge different premiums to the same individual. Building up sufficient capital resources through higher premiums would therefore be an alternative, although such premium rises might deter potential enrollees.

A limitation of our study is that we focused on 'random risk' assuming the absence of 'macro risk'. More specifically, we considered the risk for an insurer to randomly attract high

residual spenders from a risk group that – on average – has mean residual spending of zero. In practice, however, insurers are also subject to macro risk (i.e., the risk that – at the population-level – the mean residual spending for risk group $\mathcal{G}$ exceeds or falls below zero), e.g., due to inflation, a pandemic, or unforeseen price deviations concentrated in specific risk groups. For instance, pharmaceutical or technological advances in specific disease areas may disrupt the overall financial macro risk faced by the insurers. Such deviations at the macro scale might be more prevalent and severe for 'riskier' groups of individuals (i.e., individuals of risk groups with high average health expenditure). Therefore, the selection incentives due to heteroscedasticity of residual spending found in our research may be an underestimation of the level of these types of selection incentives in practice. More research is needed to extend our analysis of random risk with elements of macro risk.

Through another perspective, it could be argued that more variance in residual spending may indicate larger potential gains for insurers through efficiency, as it presents insurers more room to contract and provide health care efficiently. However, this argument is subject to the degree of efficiency already present in the insurance market itself. Additionally, such potential endeavors by insurers are again subject to risk aversion since the focus on a particular disease type may still attract unfavorable individuals from the costlier end of the particular risk group. Moreover, the fear for potential 'bad draws' may offset the appeal for the large potential gains to be made from a volatile group such as FCG38.

For the purpose of this research, we have assumed a risk equalization model that perfectly compensates for differences in mean spending across risk types. In practice, however, risk equalization models do not fully compensate for these differences in mean spending. One specific option for improvement of the Dutch model could be the implementation of additional morbidity indicators (e.g., diagnostic cost groups based on ICD-10, once this information is available for the entire Dutch population). The central point of our paper, however, is that even when the risk equalization model would perfectly compensate for differences in mean spending across risk types, some selection incentives may remain due to the substantial heteroscedasticity in residual spending. Consideration of this heteroscedasticity and its implications is therefore essential in the process of shaping and evaluating risk equalization and risk-sharing systems in health insurance markets.

3.7 APPENDIX

Table A3.1. The theoretical safety mark-up for eight risk groups under five levels of financial reserves

	Safety mark-up in a risk-rated market (€)				
Risk group	0% of average financial reserves	50% of average financial reserves	100% of average financial reserves	150% of average financial reserves	200% of average financial reserves
Entire population	10	11	12	14	16
Individuals aged 0-64	7	8	9	11	13
Individuals aged 65+	23	25	26	27	29
No morbidity flags	4	6	7	9	11
≥ 1 morbidity flags	26	27	28	30	31
Diabetics (PCG12-15)	34	35	36	38	39
Cluster of conditions (DCG10)	116	117	118	119	120
Extremely costly pharmaceutical usage (PCG38)	2,441	2,442	2,443	2,444	2,445

Note: The safety mark-up is derived through Equations 3.4–3.7 from the Methods section. For these simulations various levels of financial reserves have been applied, each reflecting a defined percentage of the average financial reserves a Dutch health insurer typically held in 2018–2019 per enrollee (€550) (Nederlandse Zorgautoriteit (NZa), 2019; 2020).

PART II

Simulating Strategies *to Mitigate Selection Incentives*

4

IMPROVING DIAGNOSIS-BASED COST GROUPS TO REDUCE INCENTIVES FOR RISK SELECTION

Based on: Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2023). Improving Diagnosis-Based Cost Groups in the Dutch Risk Equalization Model: The Effects of a New Clustering Method and Allowing for Multimorbidity. *International Journal of Health Economics and Management, 23*(2), 303-324. https://doi.org/10.1007/s10754-023-09345-0

Abstract

Health insurance markets with community-rated premiums typically use risk equalization (RE) to compensate insurers for predictable profits on people in good health and predictable losses on those with a chronic disease. Over the past decades, RE models have evolved from simple demographic models to sophisticated health-based models. Despite the improvements, however, nontrivial predictable profits and losses remain. This study examines to what extent the Dutch RE model can be further improved by redesigning one key morbidity adjuster: the Diagnosis-based Cost Groups (DCGs). This redesign includes (1) a revision of the underlying hospital diagnoses and treatments ('dxgroups'), (2) an application of a new clustering procedure, and (3) allowing multi-qualification. We combine data on health spending, risk characteristics, and hospital claims for all individuals with basic health insurance in the Netherlands in 2017 (N = 17 m) with morbidity data from general practitioners (GPs) for a subsample (N = 1.3 m). We first simulate a baseline RE model (i.e., the RE model of 2020) and then modify three important features of the DCGs. In a second step, we evaluate the effect of the modifications in terms of predictable profits and losses for subgroups of consumers that are potentially vulnerable to risk selection. While less prominent results are found for subgroups derived from the GP data, our results demonstrate substantial reductions in predictable profits and losses at the level of dxgroups and for individuals with multiple dxgroups. An important takeaway from our paper is that smart design of morbidity adjusters in RE can help mitigate selection incentives.

4.1 INTRODUCTION

Many social health insurance systems rely on regulated competition to enhance fairness and efficiency in health care financing (Ash et al., 1989; Enthoven, 1988a). Two typical regulatory features include premium-rate restrictions and risk equalization (RE). Rate restrictions – e.g., in the form of community-rating per insurance plan – protect individual affordability of health insurance for high-risk people who would otherwise be charged (very) high premiums. RE compensates insurers for the predictable variation in individual health care spending using a predefined set of risk adjusters such as the age, sex and health status of enrollees. Appropriate compensation is required to level the playing field for competing insurers, reduce the selection incentives they face, and protect the functioning of the insurance market (Rothschild & Stiglitz, 1976; Newhouse, 1996; Van de Ven & Ellis, 2000a; Layton et al., 2018c).

Although RE systems have developed from simple demographic models (adjusting only for age and sex) to sophisticated morbidity-based models – containing indicators based on inpatient diagnoses, pharmaceutical usage, and prior care spending –, recent studies have demonstrated that even advanced systems do not eliminate selection incentives (Van Kleef et al., 2018a; McGuire et al., 2020). For example, the most sophisticated RE models, applied under the Affordable Care Act in the United States (US) and in the health insurance schemes in Germany and the Netherlands, undercompensate insurers for specific, sizable groups of chronically ill consumers such as diabetics and patients with heart diseases (McGuire et al., 2020; Withagen-Koster et al., 2020).

Therefore, the aim of this paper is to diminish the persisting under/overcompensations of the Dutch RE model through the redesign of one of its key risk adjusters for morbidity, the Diagnosis-based Cost Groups (DCGs). This morbidity adjuster classifies individuals in cost groups based on inpatient and outpatient diagnoses and treatments. All selected codes relate to expected medical spending in the subsequent year and are grouped into 200 clusters of diagnoses called '*dxgroups*', based on clinical homogeneity. These dxgroups, in turn, are clustered into the aforementioned DCGs based on their respective residual spending from a prediction model for health spending that includes age, sex, and an adjuster for pharmaceutical usage (Lamers, 1998; Prinsze & Van Vliet, 2007). These DCGs are then used in the Dutch RE model that includes all risk adjusters. Although this approach is extensive, there is scope for improvement.

A potential drawback of the current DCGs is that the clustering of multiple dxgroups into one DCG inherently subjects the formed DCG to internal heterogeneity, resulting in compromised precision of compensation for specific, included diagnoses. For instance, Van Kleef et al. (2020a) found leukaemia patients to be substantially undercompensated (€3,733 on average per person per year), despite inclusion of the diagnosis 'leukaemia' in

one of the DCGs. A second potential drawback of the current DCGs is the limited ability for financial compensation of multimorbidity. Contrary to diagnosis-based risk adjusters applied in American and German RE models in which enrollees can qualify for multiple diagnostic groups (referred to as Hierarchical Conditions Categories and Hierarchical Morbidity Groups respectively), Dutch enrollees are bound to a two-layered classification system of Primary and Secondary DCGs (pDCG/sDCG) where individuals can be flagged by only one DCG per layer (Van Kleef et al., 2018a). As a result of this restriction, individuals with multimorbidity are on average undercompensated (Eijkenaar et al., 2018). Moreover, the heterogeneity within the clusters results in imperfect compensation on the level of individual dxgroups. With dxgroups derived from diagnoses and treatments, insurers have the potential to exploit knowledge on the (un)profitability of specific treatments when contracting care services. Through selective contracting, insurers could attempt to deter undesirable consumers by abstaining from contracting optimal-performing care providers for specific, unprofitable treatment groups (Layton et al., 2018b). This service-level selection undermines the functioning of the insurance market (Layton et al., 2018a).

The contribution of this paper lies in a new method for clustering the dxgroups to address the issues discussed (i.e., the heterogeneity within DCGs and the lack of multiclassification, and the selection problems that result). The essence of our approach is that we derive more accurate payment weights per dxgroup and use these for a more precise clustering into DCGs. Moreover, we allow for multi-classification: individuals appearing in a total of *N* dxgroups will qualify for *N* DCG flags and enable the corresponding insurer compensation. The expectation is that these modifications result in better compensations for chronically ill individuals, both on the level of dxgroups and on the level of disease groups. The methods we apply consist of four steps: deriving the baseline RE model as applied in the Netherlands for the year 2020, revising the dxgroups that are included in assembling the DCGs to update the 2020 model, developing and testing the updated clustering method, and finally, evaluating the effects of the new method.

As another novelty, following recent advances in evaluation metrics, this study goes beyond analyzing the effect of improving the DCGs on traditional metrics such as the R-squared (R^2) and Cumming's Prediction Measure (CPM). In addition to these metrics, this paper examines the effect on the under/overcompensation of groups that are potentially vulnerable to risk selection, given the characteristics of the Dutch health insurance system. These include the dxgroups underlying the DCGs as well as a broader set of disease groups identified through patient records from Dutch general practitioners (GPs). Under/overcompensation on the level of dxgroups indicates incentives for service-level selection (e.g., not contracting the best doctors for specific treatments) while under/overcompensation on the level of disease groups more generally indicates incentives for group-level selection (e.g., selective marketing and group discounts for supplementary insurance) (Layton et al., 2018a). By examining the effects of our DCG update on both

types of selection incentives, a broader insight of its potential strength is acquired. Clearly, the findings of our study are directly relevant for the Netherlands. Additionally, the international relevance of our analyses is to be found in the general conclusion that smart design of morbidity adjusters can help mitigate incentives for risk selection.

The remainder of this paper is organized as follows: first, background context of the history and current (2020) practice of the Dutch DCGs is provided to describe the reasoning for our study. Next, the data and methods are explained, followed by the outcomes of our analyses. The paper closes with a discussion of the conclusions and the resulting implications.

4.2 THE DUTCH DCGS: HISTORY AND CURRENT PRACTICE

4.2.1 History

Since their initial implementation in 2004, the Dutch DCGs have seen multiple revisions but have always retained the clustered composition. The original thirteen DCGs from 2004 contained 69 dxgroups, derived from a selection of ICD-coded diagnoses from inpatient hospital treatments (Prinsze & Van Vliet, 2007). However, adaptations to the funding structure of hospital treatments in 2005 changed the ICD-coding to the so-called 'diagnosis treatment combinations' (DTC), which called for a revision of the DCGs and the underlying dxgroups. Moreover, further changes to the DCGs were made in 2009 and 2012 to adhere to changes in the DTCs, seeking to capture shifting patterns in health care utilization (Van Kleef et al., 2014). The update in 2012 led to the discarding of the exclusivity of *inpatient* hospital diagnoses. By extending the DCGs to diagnoses from *outpatient* treatments by medical specialists, the predictive accuracy of the RE model increased and the undercompensations for the chronically ill declined. The expansion of DCGs resulted in a surge in prevalence: 2.7% of consumers qualified for compensation through the inpatient DCGs, while 8.5% qualified after the modification.

In a more recent update in 2015, the number of dxgroups that make up the DCGs—which had incrementally risen from 69 to 143—was increased to 189 in order to cover expansions of the benefit package and new changes in the coding system (Van Vliet et al., 2015). As a result of the increase, 9.4% of the population now qualified for DCG compensation. The most recent update of the DCG classification was performed in 2018 when the split was made between primary and secondary DCGs in an effort to compensate for comorbidity through a new maximum of two DCG flags per enrollee (before, only one was counted, i.e., the one with the highest mean residual spending). In addition, the number of dxgroups further increased from 189 to 200 (Eijkenaar et et al., 2018). With the adjuster now including fifteen pDCGs and seven sDCGs, 10.5% of the population qualifies for at least one of the DCGs.

4.2.2 Current practice

Since the DCGs are designed to predict health spending in the subsequent year through disease classification, the technique to derive an accurate estimation for the adjuster is both a clinical and statistical endeavor (Ash et al., 1989; Van Kleef et al., 2018a). First, all relevant information on hospital care is gathered to separate the primary, unambiguous diagnoses by excluding stomach complaints, for instance, and including diabetes. Second, a medical judgment is made by experts to select diagnoses that reflect a chronic condition and recurring health spending. The resulting diagnoses are then clustered to DCGs in order to minimize complexity of the RE model and simultaneously mitigate incentives for gaming (Prinsze & Van Vliet, 2007). Medical experts cluster the diagnoses into clinically more or less homogeneous dxgroups, which are subsequently clustered into DCGs by their respective residual spending from an ordinary least squares regression for health spending. The clustering is performed with Ward's hierarchical clustering method which continuously merges two clusters whose fusion results in a minimal increase in variance in mean residual spending within the newly found cluster (Lamers, 1998). This way the heterogeneity within the new clusters (DCGs) is minimized. Thus, while the clinical funnel-approach serves to select medically valid, unambiguous diagnoses that relate to future spending, the statistical counterpart is applied to derive clusters of dxgroups that maximize model fit (Ellis & Ash, 1995).

Despite the careful selection of diagnoses and the statistical procedure of clustering, undercompensations for groups of chronically ill remain. The work by Eijkenaar et al. (2018) partly addressed the heterogeneity within the DCG classification by considering comorbidity through a maximum of two DCGs (one pDCG and one sDCG) per person per year. In an effort to further improve the DCG adjuster, the authors ventured beyond the two-layered approach by deriving a model including all 200 dxgroups as separate risk adjusters (in the form of dummy variables indicating whether or not an enrollee is flagged by a dxgroup). Although this strategy improved model fit and de facto eliminated heterogeneity among dxgroups clustered in the same DCG, it was considered too complex from a practical point of view given the number of variables and the instability of payment weights for uncommon dxgroups. Therefore, we seek to reduce heterogeneity within the DCG-classification without aggravating concerns related to complexity and instability.

4.3 DATA

For this study, three datasets have been made available, which can be merged through an anonymized identification key. The first dataset includes information on spending and risk adjusters for all roughly seventeen million individuals with basic health insurance in the Netherlands in 2017. This dataset has been used for the official calibration of the Dutch RE model for 2020. The second dataset covers the roughly twenty million individual hospital

claims for specific hospital diagnoses in 2016, of which around five million are relevant for the DCG-classification of 2020. The third and final dataset includes morbidity information from electronic patient records of about 400 Dutch GP practices. In total these practices serve 1.3 million individuals. For each of these individuals, the dataset includes morbidity information according to the International Classification of Primary Care by the Wonca International Classification Committee (Het Nederlands Huisartsen Genootschap, 2024). Specifically, the dataset indicates whether or not an individual suffered from a specific chronic illness in 2016. This dataset will be used to identify disease groups for which we will calculate under/overcompensations using the alternative models. It is noteworthy that the patient records by the GP provide a valuable insight into the health status of Dutch individuals. In the Netherlands, the GP serves a gatekeeping role towards specialized (hospital) care and prescribes the use of outpatient drugs.

Since the GP dataset is essentially a subsample of the calibration dataset (N = 1.3 m versus N = 17 m), the GP subsample has been rebalanced to reflect the calibration dataset. Through a procedure of data reconciliation, the target dataset (GP data) is made to reflect the initial dataset (Dutch population) in terms of a predefined set of variables (Battaglia et al., 2009). The iterative fitting procedure determines a weighting factor for every observation within the target dataset based on the risk classes of the equalization model (discussed in 'Methods' section) After applying the weighting per individual in subsequent calculations, the target dataset reflects frequencies of risk classifications and spending that are consistent with the total population of seventeen million (for a more detailed description of the procedure, see Van Kleef et al. (2020a)). Moreover, the variables unique to the target dataset have their frequencies enlarged to an amount that reflects the complete population, based on the weighting factors.

4.4 METHODS

To modify the DCGs and to examine the effects of the modified DCGs, the methodological buildup of this study consists of four steps: (1) derive a baseline model (i.e., the actual Dutch RE model for 2020), (2) revise the dxgroups underlying the DCGs and evaluate the effect thereof on measures of explanatory power, (3) develop a new clustering method for deriving DCGs and evaluate its impact on explanatory power, and (4) evaluate the effects of the adjustments to DCGs in terms of the under/overcompensations for subgroups that are potentially vulnerable to risk selection. In the following part, the respective steps are explained in greater detail, describing the different models used in this study.

4.4.1 Step 1: Deriving a baseline model, the actual Dutch RE model of 2020

In order to set up a valuable comparison between the status quo and the proposed modifications to the model, we first calibrate a baseline model. This model (which we refer to as Model 0, or M0) mimics the official Dutch RE model of 2020. For simplicity, we solely focus on the RE model for somatic care, accounting for roughly 90 percent of spending in basic health insurance.[1] This model includes eleven risk adjuster classifications with nearly 200 risk classes in total (Van Kleef et al., 2018a; Zorginstituut Nederland, 2020). Based on the coefficients that result from an individual-level regression of health spending on these risk classes, a prediction of health spending in 2020 is made.

The risk adjusters in the baseline model include age interacted with sex, pharmacy-based cost groups (PCGs), DCGs, durable medical equipment groups, institutional status interacted with age, clusters of zip codes based on regional factors, socioeconomic status interacted with age, household size interacted with age, multiple-year high-cost groups, physiotherapy diagnosis groups, and prior spending on home care (Van Kleef et al., 2018a). Of these classifications, most are either based on demographic information or derived from information on hospital utilization. However, the role of primary care is implicitly considered in the model as well, since the PCGs are derived from drugs prescribed by GPs. So, through these PCGs, the RE model compensates for the above-average spending of chronically ill people who were not treated in a hospital in the prior year but by their GP, e.g., (most) diabetics.

4.4.2 Step 2: Revision of dxgroups underlying DCGs

After deriving the baseline model, the next steps cover the adaptations to the model. In line with the historically recurring revisions to the DCGs, the first change to the model was to update the dxgroups on both clinical and statistical grounds. The revision incorporates similar criteria as used by Ellis and Ash (1995): at least a third of the patients with the considered diagnoses need to be diagnosed in two consecutive years, the diagnoses need to be unambiguous, and the average related health spending and prevalence both need to exceed predefined thresholds. The resulting selection of diagnoses was then evaluated by a committee of clinical experts, leading to a clustering of the diagnoses into clinically meaningful dxgroups. This revision process increased the number of dxgroups from 200 to 209 – removing ten and adding nineteen –, mostly because of heterogeneity in the expenditures for diagnoses in the same clustered dxgroup that therefore needed to be separated (for the full report – in Dutch – that delves deeper into the selection of dxgroups, see Van Kleef et al. (2020b)).

1 In practice, the Dutch RE model consists of three components: one model for somatic care, one for mental care and one for out-of-pocket spending due to a mandatory deductible. See Van Kleef et al. (2018a) for more detail. The DCGs discussed in this paper are part of the model for somatic care.

To analyze the effects of the revision, we estimated an updated version of the baseline model. In this Model 1 (or M1), the DCGs of 2020 are replaced by the updated DCGs from step 2, built with the new total of 209 dxgroups. This method follows the exact same route to compile DCGs as the baseline model, as discussed in the Introduction and 'Current practice' section. Therefore, M1 has the same number of primary (15) and secondary (7) DCGs as M0 but will inevitably have different respective coefficients. We compare the explanatory power of this updated model with that of the baseline model using R^2 and CPM, which are calculated according to Equations 4.1 and 4.2, respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2}{\sum_{i=1}^{n} \left(y_i - \overline{y}\right)^2} \tag{4.1}$$

$$CPM = 1 - \frac{\sum_{i=1}^{n} \left|y_i - \hat{y}_i\right|}{\sum_{i=1}^{n} \left|y_i - \overline{y}\right|} \tag{4.2}$$

In both equations, y_i represents the actual spending for individual i, $\hat{y}_i$ the predicted value of spending for i based on the respective RE models (M0, M1, etc.) and $\overline{y}$ the average spending in the population (Layton et al., 2018a). For both measures, a value closer to 1 indicates higher explanatory power of the model. The difference between Equations 4.1 and 4.2 is that the CPM handles errors in the regression model linearly, whereas the R^2 gives a larger weight to larger errors.

4.4.3 Step 3: Developing and testing a new method of clustering dxgroups into DCGs

The third step of our empirical analysis concerns the design and evaluation of a new approach for clustering dxgroups into DCGs to mitigate the earlier described problems of heterogeneity. The new method of clustering is comprised of two phases. First, all 209 dxgroups derived from step 2 are included as separate risk adjusters in the process of calibrating the actual RE model. To be precise, the 209 dxgroups serve as explanatory variables, like age, sex, and PCGs, for the individual-level regression of health spending to obtain *coefficients*. By doing so, the respective coefficients for dxgroups are corrected for all other adjusters. In contrast, the traditional method (as used for M0 and M1, discussed in step (2) uses a partial model that corrects for age, sex, and PCGs alone to obtain *mean residual spending* for each separate dxgroup and clusters these to create the DCGs for the complete model with all adjusters. Through the new method, the respective coefficients are corrected for multimorbidity as well as potentially confounding effects of the other risk adjusters.

In the second phase of the new method, the 209 dxgroups are clustered into 26 DCGs on the basis of their estimated coefficients. By clustering dxgroups with an interval of roughly 500 euros between the lowest and highest coefficient, a manageable number of 26 DCGs are formed. These clusters are relatively stable and consist of dxgroups that

have their payment weights adjusted for multi-qualification. Thus, the primary/secondary distinction is removed from the DCGs and individuals can be assigned multiple times to multiple DCGs. As a result, health spending of individuals with multimorbidity can be more effectively compensated through RE. The model containing these new DCGs will be referred to as Model 2 (M2). To evaluate the impact of this new clustering method on the explanatory power, we compare the R^2 and CPM of M2 with those of M1.

Moreover, to further examine the effects of the new clustering method, we evaluate the impact on proximations of risk selection incentives. More specifically, we calculate the average under/overcompensations for subgroups of the population for the different models applied in this study. To analyze incentives for service-level selection, we calculate the mean financial result per dxgroup under the various models through Equation 4.3:

$$\text{Mean Financial Result } (MFR_g) = \frac{\sum_{i=1}^{N_g}(\hat{y}_i - y_i)}{N_g} \tag{4.3}$$

with *g* representing a particular dxgroup, $\hat{y}_i$ the predicted health spending for individual i in group *g*, y_i the actual spending for individual i in group *g*, and *Ng* the number of individuals in group *g*. The Mean Financial Result (MFR) that results from Equation 3 will be calculated for each of the 209 dxgroups and indicates the incentives for insurers to engage in service-level selection. More specifically, a *negative* value for the MFR for dxgroup *g* indicates incentives for insurers to select *against* users of treatments included in dxgroup *g* (e.g., by not contracting the hospitals/doctors with the best reputation regarding these treatments).

4.4.4 Step 4: Evaluating the effects of the new DCGs for selective groups

In a final empirical step of this research, the average under/overcompensations are calculated for subgroups defined by chronic illnesses in the GP data. Under/overcompensation for a subgroup of people with a specific chronic illness indicates selection incentives towards that group. Insurers in the Netherlands have several tools to select in favor of or against specific groups, e.g., through selective advertisement, group arrangements, supplementary insurance, and customer service (Van Kleef et al., 2020a; Van Veen et al., 2015). The GP data covers nearly 700 conditions classified by the ICPC. However, for this particular study, only 109 that are labeled 'chronic' will be included because the RE system is primarily designed to account for recurring expenditures related to chronic illness. The 109 conditions are selected for this study as they are – according to medical experts – unlikely to be fully recoverable from and thus relate to future health care spending (Nielen et al., 2019). A complete list of these ICPC conditions is provided in the Appendix (Table A4.1).

Similar to the use of the MFR for dxgroups to find the service-level selection incentives, we calculate the measure for each of the 109 chronic conditions in the GP data under

the various models. While the MFRs for the dxgroups specifically indicate incentives for service-level selection through the direct link with treatments, the MFRs for chronic conditions more generally indicate incentives for group-level selection (Layton et al., 2018a).

4.5 RESULTS

This section presents the results of our analyses. We first report some descriptive statistics of the datasets. Then, the measures of fit are reported for the various models, followed by the effects of the modifications to the baseline model for the dxgroups and ICPC conditions, respectively.

It is important to reiterate that we calculate the mean financial result to insurers for specific subgroups of the population to approximate selection incentives. Within that context, a *negative* value for the mean financial result implies a direct *negative* result for the insurer and incentives to select against the subgroup in question. In the other chapters of this dissertation, we calculate a value of undercompensation, where a *negative* value implies an *overcompensation* and a direct *positive* result for insurers to attract these particular individuals. This distinction is vital for the interpretation of the results of this study and the dissertation as a whole.

4.5.1 Descriptive statistics, rebalancing, and the RE model of 2020

Table 4.1 presents a selection of descriptive statistics for both the total population and the GP data. Despite the notable difference in size, the GP sample differs marginally from the population on the underlying characteristics. The individuals in the GP data are slightly older and have elevated proportions of morbidity adjuster qualifications, contributing to a higher mean spending. After applying the rebalancing process, as discussed in the Data section, the frequencies of risk classes in the GP data are identical to those of the population. Although the rebalancing process substantially reduced the difference in mean spending between the datasets, a difference of roughly five euros remained. In order to offset this leftover difference, we applied a linear correction to the spending for the individuals included in the GP sample.

Moreover, since the GP data is based on diagnoses signaled in 2016 and the population data refers to spending in 2017, individuals aged 0 in 2017 are excluded from the latter dataset. Furthermore, as consumers may, for particular reasons, cancel their health insurance at any point in the year, the expenditures on health care are annualized and weighted by the fraction of the year the individual was enrolled. Health spending incurred by individuals that were only registered in the first three months of the year, for instance, is therefore quadrupled and included in the analyses with a weight of 0.25.

Table 4.1. Descriptive statistics of the datasets included in the study

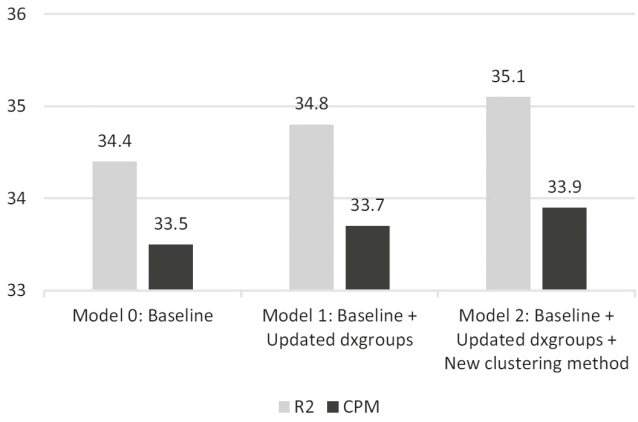
	Population	Sample (GP data)
Number of individuals	16,873,980	1,308,301
Number of insured years	16,670,015	1,299,534
Mean spending in euros per year	€2,333	€2,388
Frequencies		
Men, 1-17 years	9.8%	9.8%
Men, 18-34 years	10.5%	10.2%
Men, 35-44 years	6.0%	6.0%
Men, 45-54 years	7.6%	7.7%
Men, 55-64 years	6.8%	6.9%
Men, 65 years and older	8.8%	8.9%
Women, 1-17 years	9.4%	9.3%
Women, 18-34 years	10.3%	10.3%
Women, 35-44 years	6.2%	6.2%
Women, 45-54 years	7.6%	7.7%
Women, 55-64 years	6.8%	6.9%
Women, 65 years and older	10.4%	10.3%
Pharmacy-based cost groups	16.9%	17.4%
Primary Diagnosis-based cost groups	9.0%	9.2%
Secondary Diagnosis-based cost groups	3.9%	4.1%
Durable medical equipment cost groups	3.8%	3.8%
Physiotherapy diagnosis groups	1.9%	2.0%
Multiple-year high-cost groups	6.1%	6.3%
Prior home care spending	2.4%	2.5%
At least one morbidity adjuster	24.9%	25.4%

Note: Frequencies are calculated as a percentage of total insured years in the population and (rebalanced) sample, respectively. Individuals aged 0 in 2017 are excluded from both datasets.

4.5.2 Revision of the dxgroups underlying DCGs

The revision of dxgroups led to the removal of ten dxgroups and the inclusion of nineteen new ones. For the exact reasoning behind these modifications, we refer to the full report for a complete overview of the resulting 209 dxgroups (Van Kleef, 2020b). The effects of this seemingly minor update on the explanatory power of the model are presented in Figure 4.1. The R^2 increases by 0.4 percentage points, indicating that a larger share of variance in health spending is explained by the RE model. Moreover, the update of the dxgroups results in a 0.2 percentage point decrease of absolute differences between actual and predicted health spending, as shown by the increased CPM.

Figure 4.1. R^2 and CPM (both x100) of the three simulated RE models (N = 17m)

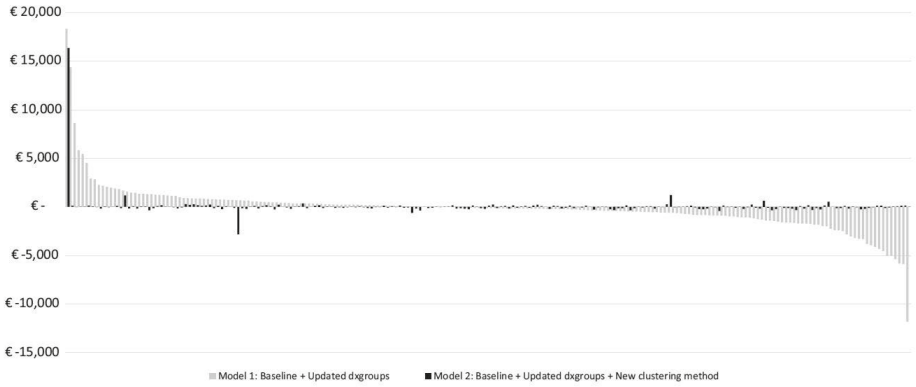


4.5.3 Developing and testing a new method for clustering dxgroups into DCGs

The update of the dxgroups and the new clustering method transformed the fifteen primary and seven secondary DCGs into 26 new separate DCGs that accommodate multi-classification. The latter means that individuals can be classified in multiple DCGs as well as multiple times in the same DCG (i.e., when they fall in multiple dxgroups included in that DCG). Moreover, the total share of the population that qualifies for compensation through DCGs increases from 10.5% to 11.4%. As a result, a larger portion of variance in health spending is explained, reflected by the increase in the measures of explanatory power: the R^2 and CPM are, respectively, 0.3 and 0.2 percentage points higher for M2 than for M1 (Figure 4.1). The R^2 and CPM of model M2—with the 209 dxgroups clustered into 26 DCGs—appeared to be nearly identical to those of the same model containing each dx-group as a separate dummy (results not shown here). This indicates that the new clustering method, while lifting complexity and stability concerns, does not cede explanatory power.

In Figure 4.2, the Mean Financial Results (MFRs) for all 209 dxgroups are projected for M1 and M2. The order of the dxgroups is based on the declining results under M1. Since the dxgroups in the figure are paired for the two models, the results for M2 do not need to follow the trend for M1. In M1, 44% of the dxgroups have a mean financial result between €500 and €-500, while in M2 over 97% fit within that range, indicating an improvement in predictive accuracy. Furthermore, of the total number of negative results for both models, 114 and 119, respectively, only 26% are between €0 and €-250 for M1, while 89% of the undercompensations in M2 are in that bracket. By expanding the lower threshold to €-500, those proportions change to 41% and 98%, indicating a noteworthy improvement.

Figure 4.2. Mean financial result for the individual dxgroups (*N* = 17m)



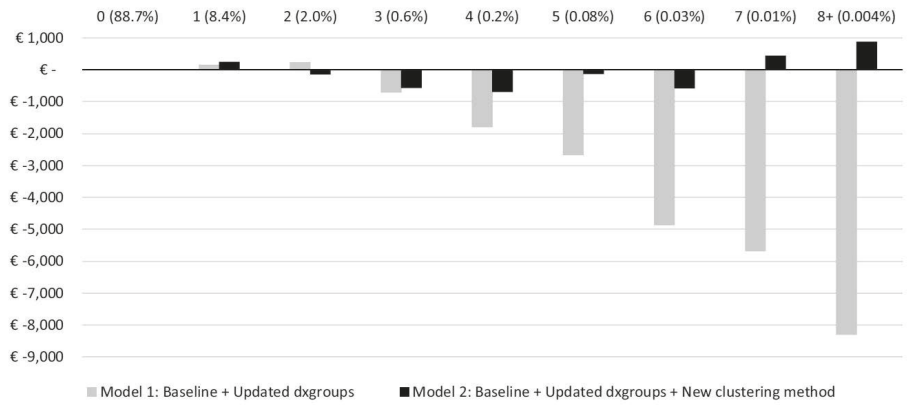
Note: Mean financial result (i.e., level of under/overcompensation) for the 209 dxgroups, paired for the two models, based on the declining rank under Model 1. The mean financial result is calculated as the mean predicted spending per person per year minus the mean actual spending per person per year.

However, a few dxgroups stand out and deviate notably from zero. This is an effect of clustering some small dxgroups with very high coefficients that are not very close to one another. This inherently leads to relatively large under/overcompensations.

The outcomes in Figure 4.2 can be summarized by the Weighted Mean Absolute Deviation (WMAD) across the 209 dxgroups. This overarching measure can be applied to reflect how capable the model is of predicting the actual mean spending for the 209 dxgroups, taking into account the relative size of the groups. Moving from M1 to M2 results in a drop of the WMAD from €653 to €113, indicating a substantial overall improvement at the level of dxgroups, in line with the findings from Figure 4.2.

Figure 4.3 presents the MFR for partitions of the population based on the number of dxgroups that individuals qualify for. The first group (from the left) represents people who are not in any dxgroup (88.7% of the population), followed by groups that are comprised of individuals that are in either one or up to eight or more dxgroups. Model 2 performs substantially better than the traditional approach of Model 1 for insured that qualify for more than two dxgroups. This indicates that allowing for multi-classification improves the outcomes of RE for people with multiple morbidities. For the larger groups of the population – those with two or fewer dxgroups – the effects are less pronounced, however.

Figure 4.3. Mean financial result for groups based on the number of dxgroups individuals qualify for (*N* = 17m)



Note: Mean financial result (i.e., level of under/overcompensation) for individuals within subgroups defined by the number of dxgroups they qualify for, within the population sample. Percentages refer to relative frequencies. Mean financial result is calculated as the mean predicted spending per person per year minus the actual spending per person per year.

4.5.4 Evaluating the effects of the new clustering for disease groups

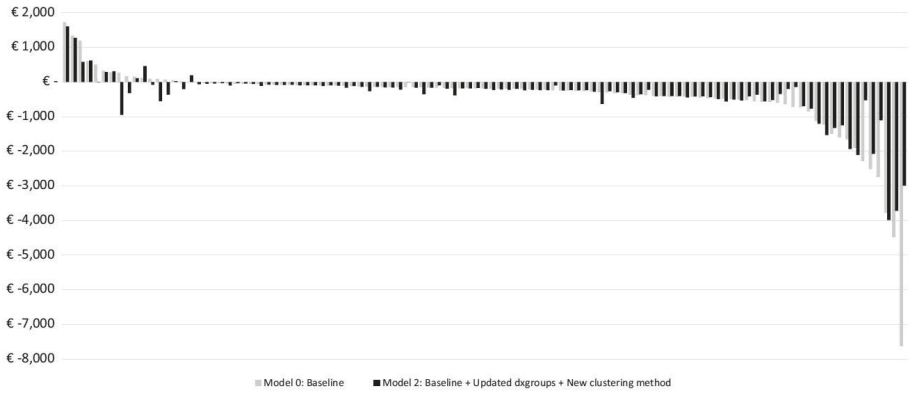
The final methodological step of this research repeats the preceding analyses performed for the 209 individual dxgroups for the 109 ICPC diagnoses extracted from the GP dataset. Thereby, the effects of the new clustering method are evaluated at the level of common chronic conditions, rather than the more specific level of hospital diagnoses. While a complete overview of the results for the 109 individual ICPC groups under both the baseline model and M2 is provided in the Appendix (Table 4.2), the key findings are reported below.

First, we made a partition between individuals that are diagnosed with at least one chronic illness on the ICPC list and those that have no such illness. The former group, the chronically ill individuals, are undercompensated by €80 on average in M0, while those without any diagnosis are overcompensated by €96. Model 2 produces an average result for these two groups of €-79 and €95, respectively, demonstrating a minimal effect of the new DCGs for this general partition of the population.

Figure 4.4 displays the MFRs derived for the 109 ICPC diagnoses for both the baseline model and Model 2. The results are paired and ordered based on the (declining) financial results of M0. The differences between the results for the two models are less obvious for these ICPC diagnoses than for the dxgroups shown in Figure 4.2. For some of the

diagnoses the MFR is closer to 0 for M2 than for M0; for other diagnoses, however, the opposite holds true. This may be explained by the fact that the new clustering method creates new clusters of dxgroups. Assume that some dxgroups were overcompensated before as a result of the inherent heterogeneity problems of the old method, and the new DCGs reduce these overcompensations. ICPC diagnoses correlated to these dxgroups are compensated to a lesser extent, producing a different financial result on average.

Figure 4.4. Mean financial result per individual ICPC diagnosis identified in the GP data (*N* = 1.3m)



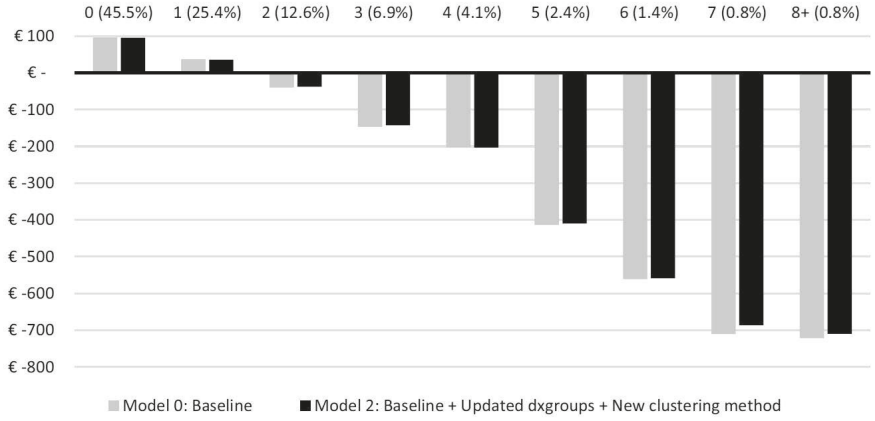
Note: Mean financial result (i.e., level of under/overcompensation) for the 109 ICPC groups, paired for the two models, based on the declining rank under Model 0. Mean financial result is calculated as the mean predicted spending per person per year minus the mean actual spending per person per year.

Nevertheless, while the baseline model produces a statistically significant MFR ($p < 0.01$) that is below zero for 38 diagnoses, model M2 reduces that number to 33. Moreover, moving from M0 to M2 results in an MFR that is statistically significantly different ($p < 0.05$) from the respective MFR produced under the baseline model for five diagnoses. Four (B73, B74, F84 & R84) of these have their undercompensations reduced by roughly 67%, tackling the problematically unprofitable aspect of these specific conditions (see the Appendix (Table 4A.1) for the details).

Similar to our earlier calculation for the 209 dxgroups in step 3, the impact of the new clustering method for the 109 ICPC diagnoses can be summarized by the WMAD: €232 for M0 and €225 for M2, a marginal three percent improvement. The limited effect on the WMAD can be partly explained by the dissimilar impact on the 109 diagnoses. Additionally, the visible changes between the models at the far right of Figure 4.4 are diagnoses with minor frequencies, whereas the hardly changed diagnoses in the center are the most prevalent.

Figure 4.5 presents the MFRs for groups of individuals based on the number of ICPC diagnoses, ranging from zero up to eight or more. In the figure, a modest difference can be observed between the bars for the subgroups of the two models. All in all, moving from the baseline model to M2 seems to have a minimal effect on the financial result of the selected groups based on the ICPC diagnoses, both in terms of multimorbidity as well as for the overall result for (not) chronically ill consumers.

Figure 4.5. Mean financial result for groups based on the number of chronic conditions identified in the GP data (*N* = 1.3m)



Note: Mean financial result (i.e., level of under/overcompensation) by number of ICPC diagnoses in the rebalanced GP sample. Percentages refer to relative frequencies. Mean financial result is calculated as the mean predicted spending per person per year minus the actual spending per person per year.

4.6 DISCUSSION

The findings of this study demonstrate how the Dutch RE model can be improved through a redesign of the DCGs. By applying a new clustering method and allowing for multiclassification, the explanatory power of the RE model at the individual level is increased. More specifically, we find an increase of the R^2 (× 100) from 34.4 to 35.1. Given the level of sophistication of the Dutch model (with multiple risk adjusters that collectively incorporate over 200 indicators), this 0.7 increase can be considered a meaningful improvement since major gains on this measure are rarely made in advanced models (Layton et al., 2018a). At the level of dxgroups the new clustering method results in substantial reductions of under/overcompensation, implying that the method better accounts for spending heterogeneity among dxgroups. For subgroups defined by ICPC diagnoses from GPs, improvements are relatively minor. The main explanation is that the new design of the

DCGs does not extend the group of people qualifying for compensation through a DCG but 'only' improves compensation among those that already did qualify. In other words, the new design moves funds *within* the group of people qualifying for a DCG but does not increase funds for this group as a whole. Although under/overcompensations reduce for some ICPC diagnoses, they increase for others, implying that the overall effect for these disease groups is limited.

All in all, it can be stated that the new approach for compiling DCGs is a successful route to improve the predictive accuracy of the Dutch model, both at the level of individuals and at the level of subgroups. In general, these improvements mean that selection incentives for insurers will be reduced since the under/overcompensations shrink (Van Vliet et al., 2015). With respect to the two sets of subgroups identified in this research (based on either the dxgroups or the ICPC diagnoses), we distinguish incentives for service-level selection and group-level selection (Layton et al, 2018a). With respect to the broader, group-level incentives derived from the ICPC diagnoses, the new DCGs produce marginal differences from the present (2021) model to stimuli for risk selection regarding disease groups. In contrast, incentives for service-level selection are more effectively addressed by our method. Since contracting of care is performed at the treatment level, the financial results for the dxgroups serve as a fitting proxy for selection incentives. By rendering the majority of dxgroups equally attractive in financial terms, as shown in Figure 4.2, our new DCG approach thus substantially reduces incentives for service-level selection in Dutch basic health insurance.

Despite the advancements made in reducing under/overcompensations for treatment groups, the Dutch RE system persistently undercompensates the overall group of chronically ill individuals, indicating that risk selection incentives towards this ambiguous group remain. A potential explanation could be that while — according to the GP patient records — 55% of Dutch individuals have a chronic condition, only 25% of the population qualifies for at least one morbidity adjuster. A substantial share of the chronically ill are not flagged by any morbidity indicator, resulting in inadequate financial compensation. The undercompensation of diabetics (€-194 under M2), for instance, may result from the disparity in its prevalence between the GP data (6.1%) and its presence in the morbidity adjusters in the RE model (less than 5%). The resulting incentives undermine the goals of efficiency and fairness in the Dutch health insurance market, accentuating the need for further improvements of the RE model. As long as groups of consumers in need of specific types of care remain predictably unprofitable, insurers face a disincentive to improve the quality of these specific types of care.

Additional research is required to detect the individuals that are chronically ill but are not found by the current RE model. While our study refines one morbidity adjuster and addresses the heterogeneity within, it is inadequate to eliminate undercompensations for all chronically ill individuals. Potentially, new risk adjusters may yield explanatory value

by including more extensive information on health or illness. Despite the fact that the GP data used in this study only covers a portion of the entire population, its usefulness should not be disregarded. Through the gatekeeping role of GPs in the Dutch health system, they possess information on the health status of individuals that may, for any particular reason, not use secondary care. Since the morbidity adjusters in the RE system are mostly based on information from hospital treatments and drug prescriptions, this could result in the mismatch between a signaled illness in the GP data and no compensation through RE. The GP data could be used for the RE model when made representative for the entire population, or efforts could be made to acquire this particular information for all individuals in Dutch basic health insurance. By doing so, the model can explicitly compensate for chronic illnesses, as indicated by the GP, rather than implicitly through hospital treatments or drug prescriptions.

Alternatively, instead of using the ICPC information for new morbidity adjusters, another promising strategy to reduce the incentives for risk selection is by applying new estimation techniques. Constrained regression, for instance, can be used to enhance the financial result for specific subgroups, such as the ICPC diagnoses (Van Kleef et al., 2020a). The model could then be calibrated to ensure a mean financial result for diabetics, or any other group, of €0 or even a profit. By using the GP information in such ways, the selection incentives towards these groups can be diminished. Another alternative may lie in shifting from the prospective nature of the Dutch RE model to a concurrent model, not based on historical information but rather based on data from the current year. Information from the current year may notably more accurately reflect the health status of an individual than historical data. This particular transition does, however, result in a larger degree of endogeneity between health expenditures incurred and compensation received. These options are interesting directions for further research.

Nevertheless, by allowing for multi-classification, the Dutch DCGs become more in line with diagnostic classifications used in RE models in the US and Germany (i.e., the Hierarchical Condition Categories and the Hierarchical Morbidity Groups, respectively). One important difference, however, remains: while the US and Germany classifications rely on direct ICD-codes, the Dutch model uses context-specific 'diagnosis treatment combinations' (DTCs). Unfortunately, these DTCs do not allow for building disease hierarchies that take disease severity into account. It is uncertain, however, if and when ICD-codes become directly available for the purpose of RE in the Netherlands; once they do, it should be possible to further refine the Dutch DCGs. Still, the new DCG approach of this study is notably effective in reducing the incentives for service-level selection, contributing to the functioning of the health insurance market. Although this study focuses on the relatively specific Dutch context, the findings and the considerations for further improvement may be valuable to regulators or health insurance systems elsewhere seeking to apply RE to counteract risk selection incentives. The overall takeaway from our analyses is that smart design of morbidity adjusters can help mitigate selection incentives.

4.7 APPENDIX

Table A4.1. Mean financial result for the 109 chronic conditions identified in the GP data (*N* = 1.3m)

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean financial result per person for Model 0 (€)	Mean financial result per person for Model 3 (€)
General and Unspecified					
A28	Limited function/disability, not otherwise specified	0.11	7,199	- 485	- 561
A79	Malignancy, not otherwise specified	0.051	17,211	** - 4,483	** - 3,720
A90	Congenital anomaly, not otherwise specified/multiple	0.19	5,262	-177	- 387
Blood, Blood Forming Organs and Immune Mechanism					
B28	Limited function/disability	0.017	7,608	- 1,498	- 1,329
B72	Hodgkin's disease/lymphoma	0.19	9,740	89	- 556
B73	Leukaemia	0.13	15,917	** - 2,285	- 530
B74	Malignant neoplasm blood, other	0.078	24,849	** - 7,627	** - 2,991
B78	Hereditary haemolytic anaemia	0.24	3,051	- 145	- 163
B79	Congenital anomaly blood/lymph, other	0.042	8,145	** - 1,597	* - 1,250
B83	Purpura/coagulation defect	0.63	6,311	- 255	- 241
B90	HIV-infection/aids	0.11	14,995	* - 573	* - 522
Digestive system					
D28	Limited function/disability	0.043	9,136	- 153	- 349
D74	Malignant neoplasm stomach	0.051	9,212	* 1,194	582
D75	Malignant neoplasm colon/rectum	0.69	9,534	- 137	- 268
D76	Malignant neoplasm pancreas	0.027	16,621	* - 2,516	- 2,076
D77	Malignant neoplasm digestive, other/ not otherwise specified	0.18	11,275	** - 1,226	** - 1,531
D81	Congenital anomaly digestive system	0.31	2,120	* - 211	* - 228
D92	Diverticular disease intestines	1.72	6,182	** - 192	** - 232
D94	Chronic enteritis/ulcerative colitis	0.81	6,761	** - 401	** - 413
D97	Cirrhosis/Liver disease, not otherwise specified	0.69	5,868	** - 571	** - 558
Eye					
F28	Limited function/disability	0.20	4,563	- 33	- 43
F81	Congenital anomaly eye, other	0.19	2,428	- 103	- 108
F83	Retinopathy	0.66	9,437	** - 559	* - 366
F84	Macular degeneration	0.62	9,036	** - 640	- 202
F91	Refractive error	3.66	2,540	- 54	- 54
F93	Glaucoma	1.54	5,255	- 33	- 49
F94	Blindness	0.19	6,020	- 325	- 324

Table A4.1. Mean financial result for the 109 chronic conditions identified in the GP data (*N* = 1.3m) (continued)

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean financial result per person for Model 0 (€)	Mean financial result per person for Model 3 (€)
Ear					
H28	Limited function/disability	0.11	4,454	329	293
H80	Congenital anomaly of ear	0.12	2,782	- 453	* - 413
H83	Otosclerosis	0.076	3,374	- 199	- 241
H84	Presbycusis	1.66	7,184	- 79	- 84
H85	Acoustic trauma	0.30	4,212	- 194	- 218
H86	Deafness	2.01	5,012	** - 216	** - 238
Cardiovascular					
K28	Limited function/disability	0.074	5,435	- 422	- 405
K73	Congenital anomaly cardiovascular	0.36	3,341	156	108
K74	Angina pectoris	2.35	7,846	** - 249	** - 231
K76	Other and chronic ischaemic heart disease	1.11	7,934	** - 382	** - 351
K77	Heart failure	1.07	14,100	** - 849	** - 768
K82	Pulmonary heart disease	0.050	17.436	- 1,126	- 1,205
K86	Hypertension uncomplicated	14.01	5,116	** - 182	** - 183
K87	Hypertension complicated	1.80	7,982	** - 468	** - 436
K90	Stroke/cerebrovascular accident	1.74	8,838	** - 366	** - 459
K91	Atherosclerosis	1.21	6,467	** - 301	** - 269
K92	Other arterial obstruction/peripheral vascular disease	2.09	7,426	** - 731	** - 700
Musculoskeletal					
L28	Limited function/disability	0.33	7,428	** - 506	** - 536
L82	Congenital anomaly musculoskeletal	1.02	2,226	- 60	* - 110
L84	Osteoarthritis of spine (any region)	1.45	6,148	** - 195	** - 229
L85	Acquired deformity of spine	0.82	3,519	** - 170	** - 192
L88	Rheumatoid arthritis and allied conditions	1.33	7,411	- 148	** - 220
L89	Osteoarthrosis of hip	2.28	7,196	** - 496	** - 509
L90	Osteoarthrosis of knee	3.32	6,783	** - 425	** - 445
L91	Osteoarthrosis, other	2.91	5,507	** - 196	** - 199
L95	Osteoporosis	2.60	6,978	** -219	** - 235
L98	Acquired deformity of limb	3.87	6,186	** -130	** - 138

Table A4.1. Mean financial result for the 109 chronic conditions identified in the GP data (*N* = 1.3m) (continued)

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean financial result per person for Model 0 (€)	Mean financial result per person for Model 3 (€)
Neurological					
N28	Limited function/disability	0.038	5,939	98	- 81
N70	Poliomyelitis/other enterovirus	0.043	6,602	- 9	198
N74	Malignant neoplasm nervous system	0.055	10,145	164	- 321
N85	Congenital anomaly neurological	0.11	9,068	273	* - 949
N86	Multiple sclerosis	0.18	11,702	- 152	5
N87	Parkinsonism/paralysis agitans	0.27	12,234	- 263	- 235
N88	Epilepsy, all types	1.01	5,508	- 52	- 34
Psychological					
P28	Limited function/disability	0.080	3,940	- 180	- 184
P70	Dementia	0.58	7,423	** 1,341	** 1,274
P72	Schizophrenia, all types	0.27	3,606	* 284	** 313
P80	Personality disorder	0.99	3,375	** - 161	** - 167
P85	Mental retardation	0.49	3,514	- 47	- 103
Respiratory					
R28	Limited function/disability	0.071	8,682	- 239	- 102
R84	Malignant neoplasm trachea/bronchus/ lung	0.21	17,065	** - 2,747	** - 1,104
R85	Malignant neoplasm respiratory, other	0.067	11,430	** - 1,656	** - 1,937
R89	Congenital anomaly respiratory	0.031	7,828	- 1,902	- 2,104
R91	Chronic bronchitis/ bronchiectasis	0.76	5,822	* - 247	* - 245
R95	Emphysema/chronic obstructive pulmonary disease	2.64	8,339	** - 422	** - 406
R96	Asthma	8.94	2,983	** - 77	** - 87
Skin					
S28	Limited function/disability	0.045	4,461	- 95	- 117
S77	Malignant neoplasm of skin	3.32	5,638	** - 169	- 97
S81	Haemangioma/lymphangioma	0.81	2,592	- 80	- 80
S83	Congenital skin anomaly, other	0.26	2,648	- 101	- 97
S87	Atopic dermatitis/eczema	9.23	2,123	* - 36	* - 35
S91	Psoriasis	2.41	4,274	** - 187	** - 169

Table A4.1. Mean financial result for the 109 chronic conditions identified in the GP data (*N* = 1.3m) (continued)

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean financial result per person for Model 0 (€)	Mean financial result per person for Model 3 (€)
Endocrine/Metabolic and Nutritional					
T28	Limited function/disability	0.0086	11,103	- 3,776	- 3,991
T71	Malignant neoplasm thyroid	0.059	5,739	121	462
T78	Thyroglossal duct cysts	0.086	3,077	- 122	- 118
T80	Congenital anomaly endocrine/metabolic	0.095	7,177	- 473	- 492
T81	Goitre, thyroid nodule including thyrotoxicosis	0.48	4,565	- 95	- 97
T86	Hypothyroidism/myxoedema	2.61	4,746	* - 92	* - 95
T90	Diabetes mellitus	6.09	7,221	** - 188	** - 194
T92	Gout	2.47	6,413	** - 405	** - 407
T93	Lipid metabolism disorder	7.75	4,550	** - 141	** - 140
Urological					
U28	Limited function/disability urinary	0.065	16,265	- 294	- 634
U75	Malignant neoplasm of kidney	0.11	11,545	* - 721	- 148
U76	Malignant neoplasm of bladder	0.27	9,757	68	- 370
U77	Malignant neoplasm urinary, other	0.017	9,692	- 107	- 164
U85	Congenital anomaly urinary tract	0.17	4,310	- 153	- 168
U88	Glomerulonephritis/nephrosis	0.14	7,638	- 382	- 228
Pregnancy, Childbearing, Family Planning					
W28	Limited function/disability	0.11	2,101	52	27
W72	Malignant neoplasm coexisting with pregnancy	0.00078	615	** 1,728	** 1,610
W76	Congenital anomaly complicating pregnancy	0.013	2,344	- 276	- 281
Female Genital					
X28	Limited function/disability	0.0067	2,563	604	623
X75	Malignant neoplasm cervix	0.22	4,555	** - 427	* - 416
X76	Malignant neoplasm breast female	1.29	6,815	19	* - 208
X77	Malignant neoplasm female genital, other	0.25	7,411	* - 524	- 413
X83	Congenital anomaly genital female	0.053	2,426	- 324	- 297
X88	Chronic cystic disease breast female	1.35	2,712	- 87	* - 97

Table A4.1. Mean financial result for the 109 chronic conditions identified in the GP data (*N* = 1.3m) (continued)

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean financial result per person for Model 0 (€)	Mean financial result per person for Model 3 (€)
Male Genital					
Y28	Limited function/disability	0.092	4,713	- 13	- 67
Y77	Malignant neoplasm prostate	0.57	8,819	** - 598	* - 347
Y78	Malignant neoplasm male genital, other	0.10	3,980	505	- 9
Y82	Hypospadias	0.087	1,849	- 52	- 45
Y84	Congenital male genital anomaly, other	0.051	1,632	- 77	- 77
Social Problems					
Z28	Limited function/disability	0.31	3,820	- 144	- 157
Weighted Mean Absolute Deviation (WMAD)				**231.6**	**224.8**

Note: Results are based on the rebalanced sample (*N* = 1.3m). Conditions for which the mean financial result in either of the two models is statistically significantly different from zero are marked with an asterisk. ** = statistically significantly different from zero with $p < 0.01$. * = statistically significantly different from zero with $0.01 < p < 0.05$. The WMAD is derived by the summation of all absolute average residuals multiplied by their respective insured years, subsequently divided by the summation of the insured years for the 109 diagnoses.

5

SUPPLEMENTING RISK ADJUSTMENT WITH HIGH-RISK POOLING TO REDUCE INCENTIVES FOR RISK SELECTION

Based on: Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2025). Supplementing Risk Adjustment With High-Risk Pooling Using Historical Data for Identifying the High Risks. *Journal of Risk and Insurance, 92*(1), 166-202. https://doi.org/10.1111/jori.12500

Abstract

Many regulated health insurance markets with community-rated premiums rely on risk adjustment (RA) to mitigate insurer incentives to risk select. However, insurers remain typically undercompensated for chronically ill enrollees. We use historical data on health spending and risk adjuster information to identify individuals undercompensated by the Dutch RA model of 2021 and find a selective group (one percent of the population) with an average annual undercompensation of €6,050. We supplement the RA model with a risk-sharing modality called high-risk pooling (HRP) to organize residual-based compensations towards insurers for the identified group to reduce the mean undercompensation to zero. The effects are evaluated on subgroups defined by chronic disease, finding a 42% reduction of their average undercompensation. Therefore, through compensating one percent of the population, the insurer incentives to select against chronically ill individuals substantially diminish. These results are compared to outlier-risk-sharing (reinsurance), proving HRP to be more effective at reducing selection incentives.

5.1 INTRODUCTION

Health insurance markets in countries like the Netherlands, Switzerland, Germany, and the United States (US) rely on regulated competition to enhance efficiency while protecting individual affordability and accessibility of coverage (Enthoven, 1988a; Glazer & McGuire, 2000; Van de Ven & Ellis, 2000a). Competition is achieved by a consumer choice of insurance plan (which leads to competition among insurers) and instruments for insurers to improve efficiency of care (such as selective contracting and alternative payment models). Typical regulatory interventions to achieve affordability and accessibility include open enrollment requirements for insurers, a standardized benefits package, premium regulations, and a risk adjustment (RA) (also known as risk equalization) scheme (Breyer et al., 2011). Premium regulations safeguard affordability for high-risk individuals (e.g., in the form of community-rating per insurance plan), while RA is complementarily designed to compensate insurers for the financial gaps between the expected spending on health services of their enrollees and the community-rated premium that is paid (Van Kleef et al., 2018b). By using a predefined set of risk adjusters such as age, sex, and indicators of health status, RA is designed to compensate insurers ex-ante for predictable variation in health spending across insurance contracts.

Through continuous advancements made over the past three decades, RA models have become increasingly accurate in predicting variation in health spending, with German, Dutch, and US models regarded as the most sophisticated. All incorporate a wide range of risk adjusters to predict individual-level spending, for instance, through clinical diagnoses and pharmaceutical usage (McGuire et al., 2021a). However, despite the quality of these models, predictable spending variation is not completely compensated, implying that predictable losses and profits for insurers remain (McGuire et al., 2021b; Van Kleef & Van Vliet, 2022b; Van Kleef et al., 2019). As a result, insurers face incentives for risk selection, which in turn may threaten the functioning of the health insurance market. Insurers may, for instance, abstain from contracting high-quality care for specific patient groups that are predicted to be unprofitable to discourage enrollment by these groups (Van de Ven et al., 2015). Recent studies have shown that health insurers indeed respond to selection incentives (Bauhoff, 2012; Carey, 2017a; Carey, 2017b; Decarolis & Guglielmo, 2017; Geruso et al., 2019; Han & Lavetti, 2017; Lavetti & Simon, 2018; Shepard, 2022; Stolper et al., 2022). In general, risk selection is undesirable since it threatens the quality, efficiency, and accessibility of the health insurance market (Glazer & McGuire, 2000; Layton et al., 2018b).

Evaluation studies of RA models share a recurring outcome: groups of chronically ill individuals are systematically undercompensated (McGuire et al., 2020; Oskam et al., 2023; Shepard, 2022; Van Kleef & Van Vliet, 2022b; Withagen-Koster, 2022; Zink & Rose, 2021). For example, Oskam et al. (2023) find that the Dutch model of 2020 leads to an average undercompensation of roughly 190 euros per person per year for enrollees with diabetes

(6.1% of the population). For enrollees diagnosed with COPD (2.6% of the population), the authors find a mean undercompensation of 420 euros and for those with hypertension (14.0% of the population), they find a mean undercompensation of about 180 euros per year. Given the substantial spending variation within these groups, not all enrollees with these conditions are in fact undercompensated. It could well be that the *mean* undercompensations are caused by a small subgroup of relatively high-risk people. Identifying these individuals is an important step towards further improvement of RA models. In a recent study, Zink and Rose (2021) applied a machine learning algorithm to identify groups of individuals that are substantially undercompensated in the US Marketplaces under the Affordable Care Act. They find that (small) groups of individuals with specific combinations of chronic conditions represent a mean undercompensation of $40,000 per person per year. In that light, the first goal of our study is to identify individuals that are predicted to be the most severely undercompensated in a different setting: the Dutch national health insurance.

However, identification of high-risk individuals alone does not solve inadequate compensation by the RA model. An intuitive option for improving compensation for the identified group of individuals is to include a (set of) risk adjuster variable(s) that explicitly flags these individuals. This approach is followed by Zink and Rose (2021), who suggest the uptake of interaction terms between disease indicators in the RA model to improve compensation for specific multimorbid individuals. In our study, this strategy seems less suitable. The reason is that (as the results of our analysis will show), spending in the prior year is a key factor in identifying high-risk individuals. For instance, the top one percent of spenders on health care in the prior year are undercompensated by roughly 5,000 euros in the current year. Although a risk adjuster for this group will reduce selection incentives, it introduces incentives for gaming. More specifically: when insurers receive additional compensation in the following year for individuals exceeding a certain cost threshold in the present year, they will be confronted with incentives to inflate spending for those approaching the threshold. Douven et al. (2015) show how such incentives are materialized in payment schemes, finding elongated treatments to reach the threshold and, as a result, an overall cost rise. Against this background, we explore an alternative solution called 'high-risk pooling' (HRP).

Proposed over two decades ago, HRP is a specific form of risk-sharing between health insurers and the regulator (or an RA fund), designed to exclusively compensate insurers for a specific group of high-risk individuals that is identified ex-ante (i.e., before the start of the contract period) (Van Barneveld et al., 1996). For this particular group, insurers receive a cost-based compensation (as a supplement to the RA payment) that reduces the undercompensation of this group. Compared to conventional risk-sharing modalities such as outlier-risk-sharing (ORS) and proportional risk-sharing, Van Barneveld et al. (2001a) found HRP to be the most effective at reducing risk selection incentives. Building on this

work, the second goal of our study is to simulate the effects of using HRP as a supplement to RA to reduce selection incentives. More specifically, we examine to what extent a supplementary cost-based compensation for the group of high-risk individuals (group H) identified in the first step of this study can help reduce selection incentives regarding a broad set of subgroups that are potential targets of risk selection by insurers. Our hypothesis is that a better compensation for group H will reduce the mean undercompensation of all groups in which H is overrepresented (e.g., disease groups like diabetes mellitus, COPD, and hypertension).

This paper proceeds as follows. In the next section, we provide a theoretical framework for the identification of high-risk individuals (our first objective) and the use of HRP to reduce selection incentives (our second objective). We then describe the data and methods used for our analyses, followed by the results. The paper ends with a section on the main findings and their implications.

5.2 THEORETICAL FRAMEWORK

This section provides a theoretical basis for the two primary aspects of our study: (1) the identification of individuals that are predicted to be undercompensated by the RA model and (2) the use of HRP as an instrument for mitigating risk selection incentives. Additionally, we explain how our paper contributes to the existing body of scientific work done on these two aspects.

5.2.1 The identification of individuals that are predicted to be undercompensated

Identifying the group of individuals with the largest predicted losses net of RA is a crucial step towards the improvement of RA models. Although state-of-the-art RA models include a broad spectrum of risk adjusters, some valuable predictors are purposely excluded from these models. Examples include spending in the previous year, diagnostic information that is considered vulnerable to gaming behavior, or health-related risk adjuster information from preceding years (Geruso & Layton, 2020). Nevertheless, any predictor not explicitly used as a risk adjuster can still help identify predictable losses net of RA.

The potential of available predictors for identifying undercompensated groups does not only depend on the predictive value of these predictors but also on the estimation method used. Like the calibration of the RA model itself, an ordinary least squares (OLS) regression could be employed to estimate individual-level profitability using various predictors. However, going beyond such transparent and well-documented methods, recent studies have successfully applied more advanced estimation techniques to better exploit comprehensive data to predict health spending (Rose, 2016; Shrestha et al.,

2018; Vimont et al., 2022). For instance, Buchner et al. (2017) and Van Veen et al. (2018) used regression trees to find interaction terms between existing risk adjusters and found additional predictive strength through the new combinations. Rose et al. (2017) have found structural undercompensation of groups using specific prescription drugs using a variety of machine learning algorithms. Zink and Rose (2021) have successfully identified individuals severely undercompensated by RA through a random forest (RF) model combining multiple conditions. With the increased access to comprehensive data on health (care utilization), machine learning algorithms can be trained to identify combinations of specific variables that predict which individuals will turn out to be substantially unprofitable. Although these complex models are often considered 'black boxes', the potential advantages in predictive accuracy are non-negligible (Irvin et al., 2020).

5.2.2 High-risk pooling

The concept of HRP was first proposed by Van Barneveld et al. (1996) and Van Barneveld et al. (1998). The essence of HRP is that insurers receive a cost-based compensation for individuals that are predicted to be undercompensated by the RA model. The ex-ante allocation of high-risk individuals to an ex-post compensation pool is what separates the modality from conventional risk-sharing modalities (Van de Ven & Ellis, 2000a). In a simulation study, Van Barneveld et al. (2001a) evaluated the effects on selection incentives when using either proportional risk-sharing (a cost-based compensation equal to a defined proportion of spending), ORS (a cost-based compensation for a proportion of individual-level spending above a defined threshold), or high-risk pooling (HRP), all with a similar risk-sharing budget. The authors found the latter to be the most effective at reducing undercompensations for groups of interest. The explanation is that HRP directs payments to where they are most needed: groups of individuals that are *predicted* to be undercompensated the most and are therefore subject to the strongest selection incentives. Moreover, another favorable aspect of HRP lies in the theory that insurers will likely have access to predictive information on the profitability of (types of) individuals and, as a result, will face selection incentives. By compensating insurers for individuals that are identified as high-risk, such selection incentives will likely be reduced.

While risk-sharing reduces selection incentives, a drawback is that incentives for efficiency are simultaneously weakened (Brammli-Greenberg et al., 2019; Geruso & McGuire, 2016; McGuire & Van Kleef, 2018b; Newhouse, 1996). Considering this tradeoff, targeting risk-sharing funds to groups that are predicted to be undercompensated might be preferred over more generic risk-sharing modalities (Van Barneveld et al., 1996; Van Barneveld et al., 1998). The approach has been empirically tested and proven to be an effective strategy in recent studies but has – to the best of our knowledge – not been implemented in any health insurance system (McGuire et al., 2020; Withagen-Koster, 2022). Therefore, this study seeks to contribute by demonstrating the potential of this targeted form of risk-sharing as a supplement to RA.

The effectiveness of HRP hinges on its design, which comes with three important choices: Which individuals are allocated to the compensation pool? How is the spending of the selected individuals compensated? And how are those compensations financed? The number of individuals that will eventually be assigned to the pool depends on the preferences of the regulator since increasing the size of the pool, and consequently the budget used for risk-sharing, reduces incentives for both risk selection and cost-containment. Moreover, the adequate identification of high-risk individuals is crucial and depends on the data and estimation methods used. Regarding the compensations, Van Barneveld et al. (2001a) suggest compensating either all expenditures of the pooled individuals or (a percentage of) their expenditures above a defined threshold. However, recent studies have proposed to target risk-sharing payments to outliers in *residual* spending, that is, spending net of RA payments. Such residual-based payments avoid 'double payments' and compensate for high spending net of RA rather than high spending per se (McGuire et al., 2021a; Schillo et al., 2016). With regard to the financing of compensations, various options exist. First, the pool could be financed either externally (i.e., extra funding towards the RA system) or internally (i.e., redistribution of the funds already in the RA system, keeping the total budget intact). In the case of internal financing, the HRP compensations can be financed by a deduction of the RA payment in the form of either a flat rate or a proportion of the RA payments for the individuals outside of the high-risk pool (Van Barneveld et al., 2001a).

5.2.3 Scientific contribution

Taken individually, the two objectives of this paper are not new. The identification of undercompensated individuals has been a topic of research since RA models have been implemented and the concept of HRP was introduced decades ago by Van Barneveld et al. (1996). The key novelty of this paper is to be found in the combination of these two streams. More specifically, we show how HRP can help to improve compensation for people that are predicted to be the most undercompensated using a broad spectrum of historical data. Our empirical simulations show that HRP does not only reduce undercompensation of the group assigned to the high-risk pool (group H), but also the undercompensation of many disease groups in which H is overrepresented. This leads us to the conclusion that HRP is a powerful tool for mitigating the selection incentives that remain under state-of-the-art RA models. This finding is particularly relevant since the possibilities for improving RA models by implementing new risk adjusters seem to be limited.

Our application of HRP shows some similarity with the work by Withagen-Koster (2022), who also simulates the effects of HRP as a supplement to the Dutch RA model. An important difference is that we rely on a broad spectrum of administrative data to identify the 'high-risks', while Withagen-Koster uses survey information for this purpose. A comparison of our findings with those of Withagen-Koster reveals that administrative information – especially prior spending – is more predictive of residual spending (i.e., spending

net of RA) than survey information. Another important difference between our work and that of Withagen-Koster is that we compare the outcomes of HRP with a more traditional and commonly applied form of risk-sharing: ORS (sometimes also referred to as reinsurance). Our findings show that – for a given risk-sharing budget – HRP is more effective in reducing selection incentives for groups of chronically ill than outlier-risk-sharing (ORS). This leads us to the following policy recommendation for mitigating selection incentives: consider replacing existing modalities of outlier risk-sharing by HRP. As an interesting byproduct of our analysis, we also find that adjusting the financing structure of existing modalities of ORS might help reduce existing selection incentives.

5.3 DATA AND METHODS

Our empirical analyses are based on data from the Dutch national health insurance. In this scheme, insurers compete on the price and quality of insurance plans within a regulatory framework to safeguard individual affordability and accessibility of coverage. All individuals living or working in the Netherlands are mandated to buy basic health insurance, providing a sizeable (roughly 17 million) cohort for our study. Other regulatory features of the Dutch scheme include open enrollment, a standardized benefits package, and a requirement of community-rating per insurance plan. Below, we describe the data (Section 5.3.1) and methods (Section 5.3.2) for our empirical analyses.

5.3.1 Data

As will be described in more detail in the Methods section, our empirical analyses roughly consist of three steps: simulating the Dutch RA model, identifying high-risk individuals, and simulating and evaluating HRP as a supplement to RA. Below we describe the datasets used for these steps.

5.3.1.1 Data for simulating the Dutch risk adjustment model for 2021

The first set of data covers information on health spending for all individuals with basic health insurance in the Netherlands in 2018 and their individual risk adjuster information.[1] This dataset has been used for the calibration of the actual RA model for 2021 and will be used to replicate it.[2] The model contains eleven risk adjusters: age interacted with sex, socioeconomic status and source of income both interacted with age, region, household size interacted with age, pharmacy-based cost groups, diagnosis-based cost groups,

1 Since RA payments must be estimated before the start of the contract year of insurance and information on individual-level health spending suffers from a natural delay of claims, information derived from earlier years is used in the calibration of the model. Therefore, for the model of 2021, information on spending from 2018 is used. This data from 2018, however, has been made representative for the year 2021 (e.g., in terms of health care services covered by the benefits package).

2 In the Dutch RA system, three separate models are used: one for somatic, curative care, one for curative mental care and one for the out-of-pocket expenditures under the mandatory deductible (€385). This study exclusively focuses on the somatic model, which covers roughly 90% of total spending in Dutch basic health insurance.

durable medical equipment cost groups, physiotherapy diagnosis groups, multiple-year high-cost groups, and prior spending on home care. The latter six risk adjusters are regarded as the most direct indicators of health status; Table 5.1 provides a brief overview of how these adjusters are operationalized. For further details about the Dutch RA model and the Dutch system in general, see Van Kleef et al. (2018a).

Table 5.1. Operationalization of six important risk adjusters in the Dutch risk adjustment model of 2021

Risk adjuster	Operationalization
Pharmacy-based cost groups (PCG)	38 classes based on individual usage of pharmaceuticals prescribed by their General Practitioner (GP) in year *t-1*. Individuals qualify for a PCG when they exceed a predefined threshold of daily doses of linked pharmaceuticals.
Diagnosis-based cost groups (DCG)	26 classes based on clusters of specific hospital inpatient and outpatient diagnoses in year *t-1*.
Durable medical equipment cost groups (DMECG)	14 classes based on the use of durable medical equipment in year *t-1* related to specific chronic conditions.
Physiotherapy diagnosis groups (PDG)	4 classes based on specific diagnoses from physiotherapy visits in year *t-1*.
Multiple-year high-cost groups	8 classes based on high health care spending on somatic care in years *t-1*, *t-2* and *t-3*. For instance, spending three times in the top 0.5% is one of the classes included.
Prior spending on home care	9 classes based on cumulative spending on home care in years *t-1*, *t-2* and *t-3*.

5.3.1.2 Data for identifying high-risk individuals net of risk adjustment

The second dataset includes data to identify individuals that are predicted to be undercompensated by the RA model. Since the information provided in section 5.3.1.1 is used for the calibration of the RA model itself, it holds little extra value for our identification strategy: the model is designed to project a net residual spending of zero for the separate classes of the included risk adjusters. To identify high-risk individuals, we supplement the data described in section 5.3.1.1 with two types of information: (1) the individual-level health spending in the prior year (*t-1*), and (2) the risk adjuster information used in the prior year.[3] The information on prior-year spending will likely yield predictive power with regard to spending in year t and may contribute to predicting residual spending in year t. The risk adjuster information of the prior year covers the variables presented in Table 5.1 but based on year *t-2*, rather than year *t-1*.[4] So, while the pharmacy-based cost groups provided in section 5.3.1.1, for instance, are based on pharmaceutical usage in year *t-1*, the newly added data provide classes that are based on pharmaceutical usage in year *t-2*. The historical risk adjuster information may hold predictive power to identify groups of individuals that are left with residual spending net of RA payments in year *t*. Through

3 Health spending in year *t-1* is used as the true costs incurred, rather than using annualized costs.

4 Note that the final two risk adjusters from Table 5.1. are based on multiple years in the past. For the data described in section 5.3.1.2, the years mentioned for the two adjusters shift from '*t-1, t-2* and *t-3*' to '*t-2, t-3* and *t-4*'.

an anonymized identification key, the described information can be merged to predict undercompensations in year *t* for each individual.

5.3.1.3 Data for evaluating the impact of risk-sharing on selection incentives

The third dataset includes information to evaluate the effect of HRP on selection incentives. This dataset includes diagnostic information from 2017 obtained from the electronic patient records of about 1.5m patients of general practitioners (GPs).[5] Diagnoses are coded according to the International Classification of Primary Care by the Wonca International Classification Committee (Het Nederlands Huisartsen Genootschap, 2024). 109 diseases in the dataset have been labeled 'chronic'. Using this dataset, we derived subgroups that are potentially vulnerable to risk selection by insurers. As will be described in the Methods section, we evaluate the effects of HRP on selection incentives by evaluating the predictable profits and losses for these groups before and after HRP.

Since the GP data is available for only a portion of the population, it has been rebalanced to reflect the Dutch population with basic health insurance (section 5.3.1.1). Through an iterative fitting procedure, a weighting factor is calculated for every observation of the smaller sample, based on the risk classes of the RA model. By using those weighting factors in the calculations, the frequencies of the GP subsample represent that of the overall population (Battaglia et al., 2009; Van Kleef et al., 2020a).[6] The prevalence of the 109 diseases is thus extrapolated to the entire Dutch population, facilitating an evaluation on a population scale for the respective conditions. Through a similar identification key as used in section 5.3.1.2, this final dataset can be merged with the others.

5.3.2 Methods

The study design comprises five steps, discussed in the following sections.

5.3.2.1 Step 1: Simulating the 2021 risk adjustment model

First, we replicate the Dutch RA model of 2021 to calculate the RA payments for each individual based on the appropriate risk adjuster information (see section 5.3.1.1 for an overview of the model). Those payments are then subtracted from the individual-level expenditures to derive 'residual' spending, which in turn serves as the baseline for our

5 This dataset is of particular value to our analyses because of the gatekeeping role of GPs in the Dutch health care system. The GP serves as the first point of contact for individuals requiring health services and possesses information on the individual-level health status. This information provides the most complete insight into the health of individuals: if a person suffers from diabetes, it is highly likely that the electronic patient record of that patient (as used by the GP) makes mention of diabetes; and if the electronic patient record makes no mention of diabetes it is highly likely that the person does not suffer from diabetes. By contrast, hospital claims cannot confirm the absence of such a condition since not all individuals will use hospital services for their condition. From the GP-dataset, subgroups based on the presence/absence of 109 chronic conditions can thus be used for quantifying the risk selection incentives with respect to (un)healthy individuals. In particular the chronically ill subgroups could be subject to risk selection incentives, because of their chronic need for health care services (Van Kleef et al., 2019).

6 Individuals aged zero in the RA data (year *t*) are excluded from the analyses because they are not represented in the GP subsample, which is from year *t-1*. Therefore, the rebalancing procedure does not entirely accurately reflect the overall population since it excludes newborns in year *t*.

analyses. Residual spending indicates to what extent (groups of) individuals are undercompensated (or overcompensated) by the model.

The calibration of the RA model is summarized through Equation 5.1:

$$f\left(risk\ adjusters\right) + \ e_1 \tag{5.1}$$

with OLS regression f applied to predict individual-level spending y, using 'risk adjusters' consisting of 219 dummy variables based on the information described in Section 5.3.1.1. The individual-level residual spending net of RA, RS_i, is then derived from Equation 5.2:

$$RS_i = \ y_i - \ \hat{y}_i \tag{5.2}$$

with y_i the actual spending for individual i (as observed in the dataset on which the RA model is estimated) and $\hat{y}_i$ the predicted spending for i according to the RA model (Equation 5.1).

5.3.2.2 Step 2: Identifying individuals with the highest predicted spending net of risk adjustment

In the next step, we apply three different techniques to identify the top X% of individuals predicted to be undercompensated by the RA model (for X = 1, 2, 3, 4, and 5). The first technique is a pragmatic approach and simply takes individual-level health spending of the prior year (y_{t-1}) as a basis for assignment to the top X%: the top X% spenders in year *t-1* are the identified high-risk individuals. The hypothesis is that the level of spending in year *t-1* is positively correlated with the level of undercompensation in year *t*.

The two other techniques to identify high-risk individuals use supplementary information to predict undercompensations, in addition to the information of prior-year health spending. Risk adjuster information of the prior year (section 5.3.1.2) is added to provide a richer range of predictive information to indicate the top of undercompensated individuals. Health spending of the prior year is also added to the model - differently from the first method: in the form of dummy variables for percentiles (or smaller fractions such as the 99.99th percentile). A continuous variable for prior-year health spending could be used, but with a disproportionate share located in the end of the distribution, the nonlinear distribution of spending is better accommodated in the estimation through dummy variables (see Table 5A.1 in the Appendix for more information). Both the second and third techniques to identify high-risk individuals thus use multiple variables to predict undercompensations, rather than prior-year health spending alone.

Both techniques can be summarized through Equation 5.3, combining the residual spending (RS) produced through Equations 5.1 and 5.2 with the supplemented historical information to calculate a prediction of the residual spending in year *t* through:

$$RS = h\,(predictors) + \ e_2 \tag{5.3}$$

with 'predictors' consisting of a set of 238 dummy variables based on information not included as a risk adjustor in the RA model for year *t* (i.e., spending from year *t-1*, socioeconomic information from year *t-1* and morbidity indicators based on health care utilization in year *t-2* and before). Using Equation 5.3, we obtain a prediction of residual spending ($\widehat{RS}$) for each individual. The top X% of these predictions indicates the individuals at the top of predicted undercompensations. Crucially, the operationalization of h is what separates the two estimation techniques: for model 2, h resembles an OLS regression, and for model 3, h resembles a random forest (RF) algorithm.[7] Although we focus on OLS and RF, h can theoretically be replaced by any other estimation technique that might be considered interesting for a specific setting.

To evaluate the three techniques for identifying high-risk individuals, we compare the three groups identified by the three models, for a given top X%. The model that isolates the high-risk group with the largest mean residual spending (RS) under the 2021 RA model (section 5.3.2.1) will be considered the most effective. Since the goal of our HRP modality is to compensate for predicted losses, we allocate to a high-risk pool the individuals that have the largest predicted residual spending.

Last, to avoid problems related to overfitting, we use a split sample methodology for all three techniques: a division of the dataset into a training and a test sample (50% of the dataset is randomly allocated to either subset).[8] The training sample is used to estimate the parameters for the models, while the test sample is used to evaluate the models. For the simulation of the RA model for 2021, for instance, we ran the OLS regression (Equation 5.1) on the training sample to derive the payment weights per risk adjuster and used those to estimate the RA payments for the individuals in the test sample. For the models based on Equation 5.3, we used a similar approach.

7 The RF procedure creates multiple decision trees from randomly selected explanatory variables applied to find the best estimates of residual spending using the information supplied (Breiman, 2001). The algorithm combines various regression trees of the variables to create 'ensembles' to facilitate better and more generalizable predictions than simple regression trees (Roy & Larocque, 2012). Moreover, the combination of different regression trees reduces the problem of 'overfitting': the model fitting too closely to the role of chance within the dataset and working less effective on new data (Dietterich, 2000; Hawkins, 2004). The hypothesis is that the algorithm produces a better predictive performance at the sacrifice of decreased transparency, creating a tradeoff between the two. The RF analysis is performed in an R environment with package 'ranger' (Wright et al., 2019).

8 To facilitate our proposed evaluation of the impact of risk sharing modalities, all individuals from the GP subsample are allocated to the test sample.

5.3.2.3 Step 3: Simulating high-risk pooling as a supplement to risk adjustment

In this next step, the individuals identified in step 2 are assigned to a high-risk pool for compensation through HRP. To maximize the potential benefits of using HRP as a supplement to RA, the allocation to the pool is determined by the most effective technique found in step 2.

The aim of HRP is to eliminate the mean undercompensation of the identified group, which implies that the total compensation to the pool equals the sum of residual spending in the pool. A question, however, is how to allocate the compensation to individuals in the pool. In line with Schillo et al. (2016), we base the allocation on residual spending. More specifically, individuals within the pool with residual spending above threshold T receive a compensation equal to their residual spending above T. Those within the pool with residual spending below the threshold will not receive nor fund any (additional) compensation. Through iterative testing, we found the value for T that reduces the average undercompensation of the high-risk pool to zero. The total budget required is therefore equal to the average undercompensation multiplied by the size of the identified pool.

Next, the financing of the compensations must be determined. In line with Van Barneveld et al. (1996) we finance the HRP by deducting a flat contribution from all individuals *not assigned to the high-risk pool*. Note that this flat contribution eliminates the mean *overcompensation* that exists for this group under the current RA model (the two groups of individuals combine to a residual spending of €0). Reducing overcompensations alters the incentives for insurers to attract these individuals. However, provided the flat contribution is deducted from all individuals outside of the identified pool, the impact is spread over a significant portion of the population.

5.3.2.4 Step 4: Evaluating the effects of high-risk pooling on selection incentives

To evaluate the impact of our risk-sharing design, the selection incentives are quantified, both pre-HRP and post-HRP. To quantify the selection incentives faced by insurers towards specific groups of individuals, we assess the undercompensation or overcompensation[9] of specific subgroups (reflecting the mean (un)profitability of individuals from that subgroup for insurers) (Van Veen et al., 2015). Thus, to assess the level of these incentives, we calculate the mean undercompensation for every subgroup, *g*:

$$Undercompensation = \overline{y}_g - \widehat{\overline{y}}_g, \tag{5.4}$$

where the mean undercompensation for a subgroup is derived from the mean spending of individuals within the subgroup and their mean level of predicted spending (i.e., RA

9 Note that undercompensations is similar to residual spending. However, to provide a simpler separation between the design of the risk sharing modality and the evaluation of its effects, for the latter we refer to the undercompensation of subgroups of the population.

payments). However, in the presence of a risk-sharing scheme, the result is not exclusively determined by the actual spending in year t and the RA payments received. Depending on whether the individual qualifies for additional compensation through the active risk-sharing modality, the individual-level result is affected by risk-sharing. To evaluate the impact of a risk-sharing modality on selection incentives, we calculate undercompensations or overcompensations for relevant subgroups of chronically ill people identified in the GP data.

5.3.2.5 Step 5: Comparing high-risk pooling to outlier risk-sharing

In a last section of our analysis, we compare the outcomes of HRP with those of a more conventional risk-sharing modality: ORS (also known as reinsurance). While HRP is targeted towards an ex-ante group of high-risk people, ORS compensates for the expenditure above a predefined threshold for any individual exceeding that threshold. In other words, with HRP people with the highest *predicted losses* are assigned to the pool, while with ORS those with the *highest spending* are assigned to the pool. Our comparison of HRP with ORS demonstrates the added value of using a targeted approach with assignment based on predicted losses relative to the conventional approach with assignment based on actual spending.

Two dimensions are relevant for the comparison of HRP and ORS modalities: the selection incentives and the efficiency incentives. In our analyses, we try to balance the efficiency incentives between the modalities by using the same compensation budget for both. Through a similar iterative procedure described to find threshold T for HRP, we determine threshold T^* for ORS such that the sum of cost-based compensations for ORS equals that of HRP. Between the modalities, the financing of compensations is similar: insurers pay a flat contribution for each enrollee not assigned to the pool (with the sum of contributions to the pool being equal to the sum of compensations from the pool). As a result, the modalities differ on two levels: the assignment of individuals to the pool (based on predicted losses vs. based on actual spending) and the allocation of compensations within the pool (100% of residual spending above a threshold vs. 100% of actual spending above a threshold).

5.4 RESULTS

This section presents the results of our analyses, including descriptive statistics for the datasets involved (section 5.4.1), a comparison of the three techniques to identify individuals that are predicted to be most undercompensated (section 5.4.2) and the design and evaluation of the risk-sharing modalities (sections 5.4.3 and 5.4.4 for HRP and ORS, respectively).

Table 5.2. Descriptive statistics from the datasets used for this study

	Population	GP data (rebalanced)	Training sample	Test sample
Number of insured	17,256,295	1,518,409	8,628,148	8,628,147
Sum of insured years	16,951,700	16,782,220	8,487,093	8,484,607
Mean spending year *t*	€2,408	€2,383	€2,407	€2,407
Mean predicted spending year *t*	€2,408	€2,387	€2,407	€2,407
Mean residual spending year *t*	€0	€ - 4	€0	€0
Frequencies				
Women	50.5%	50.5%	50.5%	50.5%
Zero years of age	1.0%	0%	1.1%	0.9%
Ages 1 to 17	18.7%	18.9%	18.7%	18.8%
Ages 18 to 34	20.8%	21.0%	20.8%	20.8%
Ages 35 to 44	11.9%	12.0%	11.9%	11.9%
Ages 45 to 54	14.8%	14.9%	14.7%	14.8%
Ages 55 to 64	13.5%	13.7%	13.5%	13.6%
Ages 65 to 74	11.1%	11.2%	11.2%	11.1%
Ages 75 and older	8.1%	8.2%	8.2%	8.1%
PCGs > 0	16.8%	16.9%	16.7%	16.8%
DCGs > 0	11.4%	11.5%	11.4%	11.4%
DMECGs > 0	4.5%	4.5%	4.5%	4.5%
PDGs > 0	2.0%	2.0%	2.0%	2.0%
Multiple-year high-costs	6.1%	6.2%	6.1%	6.1%
Prior spending on home care	2.4%	2.4%	2.4%	2.4%
At least in one of the above morbidity classes	25.2%	25.4%	25.2%	25.2%

Note: For the 'morbidity classes' (in the seven bottom lines) additional explanation is provided in Section 5.3.1.1.

5.4.1 Descriptive statistics

Table 5.2 presents an overview, with respect to individual-level health spending, age, sex and morbidity status – as applied in the RA model –, of the datasets used in the analyses. The second column and third column facilitate a comparison between the overall population and the GP data – after the latter has been rebalanced to resemble the former. Since children below the age of one are excluded from the subsample (as explained in 5.3.1.3), the prevalence of their corresponding age group is low, and the relative frequencies of the other age groups are offset. Moreover, as a result of excluding the relatively high health costs of newborns, the overall mean spending for the rebalanced subsample is lower.[10]

As discussed in section 5.3.2.2, the total population has been evenly split into a training and test sample for our analyses. These samples have been randomly assembled, but all

10 As a result of the exclusion of newborns, the RA model does not produce a mean residual spending of zero for the GP data. Through a small correction we return the average to €0.

individuals from the GP data section 5.3.1.3) are assigned to the test sample. The samples are relatively similar in terms of spending and frequencies of the considered subgroups in Table 5.2, indicating a fair split of the baseline dataset.

5.4.2 Identifying individuals predicted to be the most undercompensated

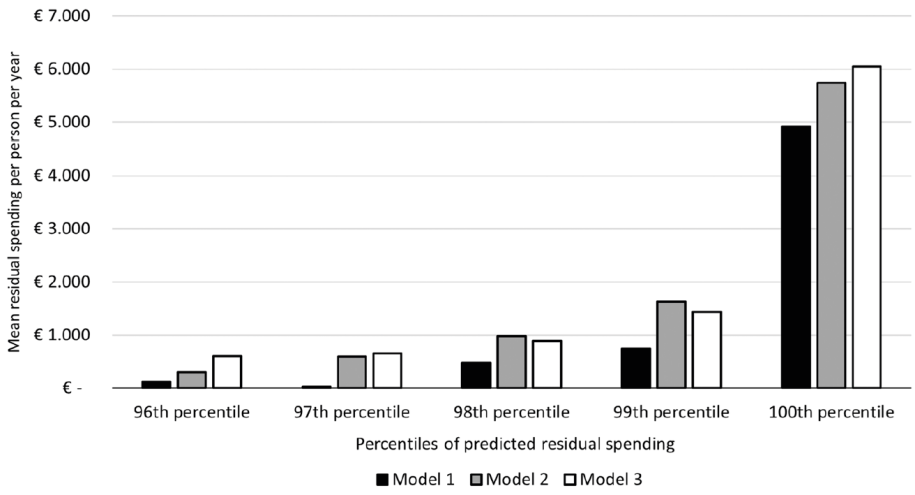
Using the training and test samples, we used three techniques to identify the most severely undercompensated individuals. The first is a model based on prior-year health spending alone, and the other two use additional historical risk adjuster information but use different estimation methods (OLS vs. RF). We do not rule out better performances of similar or different machine learning algorithms, but our analyses demonstrate only minor gains to predictive strength when recalibrating the RF model used.[11]

To determine which of the techniques is most successful in identifying the undercompensated individuals, the respective mean residual spending of the top five percentiles of individuals of the separate models is shown in Figure 5.1. Clearly, the results for Model 1 show that prior-year spending is predictive of residual spending: the mean residual spending for the top one percent of spenders in the prior year is substantial (€4,918). However, supplementing that information with historical risk adjusters results in a more selective group. The mean residual spending of all considered groups is greater for Model 2 and Model 3 than for Model 1. However, when considering groups larger than the top one percent (or the 100th percentile), the advantage of Model 2 and 3 over Model 1 diminishes as the visual differences between the bars reduce. Since the aim of this analysis is to identify the group of individuals that are most severely undercompensated, Model 3 is selected for the further steps in our simulation. With the highest percentile of individuals demonstrating the largest mean residual spending and the notable decrease in the average when expanding the group (e.g., by including the 99th percentile), we focus our next analyses on the 100th percentile of individuals identified by Model 3. (Note that in practice it is up to policymakers to choose an appropriate group size in light of payment system objectives.)

In an additional analysis, we compared the identified group (i.e., the top one percent of predicted residual spending identified by Model 3) with the rest of the population in the test sample. The outcomes are included in Table A5.3 in the Appendix. Unsurprisingly, the top one percent is considerably older and sicker. Nearly all the individuals in the top one percent have a positive score on at least one morbidity adjuster, yet still, the group is

11 In the RF algorithm we set the hyperparameters to grow 200 independent trees with a minimal sample size of 150 individuals per end node. These are set based on an iterative process of optimization. Table 5A.2 in the Appendix provides an overview of multiple combinations of minimal node size and number of independent trees grown per algorithm and the mean undercompensation for the top percentiles of identified individuals. Since the goal of the analysis is to identify a group of individuals that has the highest average undercompensation, the optimization of the hyperparameters are exclusively based on that criterion.

Figure 5.1. Mean residual spending under the Dutch RA model of 2021 for the top five percentiles of predicted residual spending by three models.



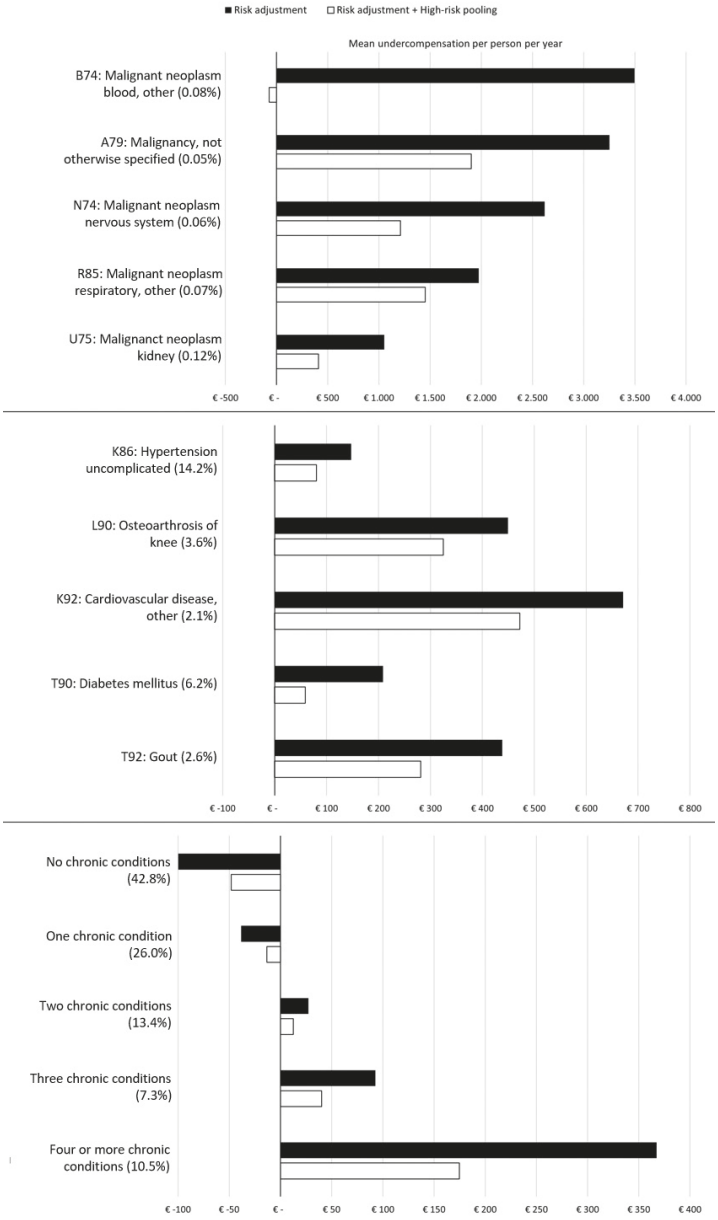
Note: The mean residual spending (y-axis) is calculated through Equation 5.4 (section 5.3.2.4) for the subgroups on the x-axis. These subgroups are derived from the predicted residual spending (Equation 5.3 in section 5.3.2.2) of the three separate models. The first model is based on health spending in the past year. The second model is an OLS regression on prior-year health spending and historical risk adjuster information. The third model is an RF algorithm based on the same information as Model 2. The three models were estimated on the training sample and then used to predict residual spending on the test sample. The predictions on the test sample were used to identify the groups on the x-axis for which we then calculated the mean residual spending under the Dutch RA model of 2021.

inadequately compensated for by the RA model. We ran the comparison for all morbidity classes in the RA model and found that the relative prevalence in all these classes was higher for the top one percent.

5.4.3 The design of high-risk pooling and evaluating its effects on risk selection

After identifying the individuals with the highest predicted undercompensation, we shifted our focus towards a strategy for reducing these undercompensations: HRP. As described in the Methods section, our goal was to eliminate the undercompensation for the identified group. To do so, we compensated the residual spending of the identified individuals above threshold T. As explained in section 5.3.2.3, the total budget that is required is equal to the average undercompensation multiplied by the size of the identified pool. Found through iterative testing, the average undercompensation of the identified pool can be reduced to zero by implementing a threshold of €18,611. Undercompensations exceeding this threshold will be reduced to this value, while those below remain

Figure 5.2. Mean undercompensation per person per year for specific subgroups under the Dutch risk adjustment model 2021 before and after high-risk pooling.



Note: Prevalence of the subgroup is shown between brackets. Results are based on the test sample. [Panel 1] Top 5 most severely significantly ($p < 0.01$) undercompensated diagnoses under RA model 2021 and their undercompensation before and after HRP. [Panel 2] Top 5 diagnoses with the largest macro undercompensations under RA model 2021 and their undercompensation before and after HRP. [Panel 3] Individuals grouped by the number of chronic diseases diagnosed by the GP and their undercompensation before and after HRP. The scale of the x-axis differs per panel.

unchanged. This strategy omits compensations to individuals that are overcompensated already and addresses the largest undercompensations first. To finance these cost-based compensations (2.6% of the total spending in the dataset), the RA payments of the excluded 99% of the test sample are deducted by a flat contribution of €61 (equal to the average overcompensation of that group under the Dutch RA model of 2021).

As a next step, we simulated the effect of HRP for selective subgroups derived from the GP data. More specifically, we calculated the mean undercompensation (before and after HRP) for 109 chronic illnesses. The full list of these illnesses is included in the Appendix (Table 5A.4); Figure 5.2 presents the outcomes for three selections.

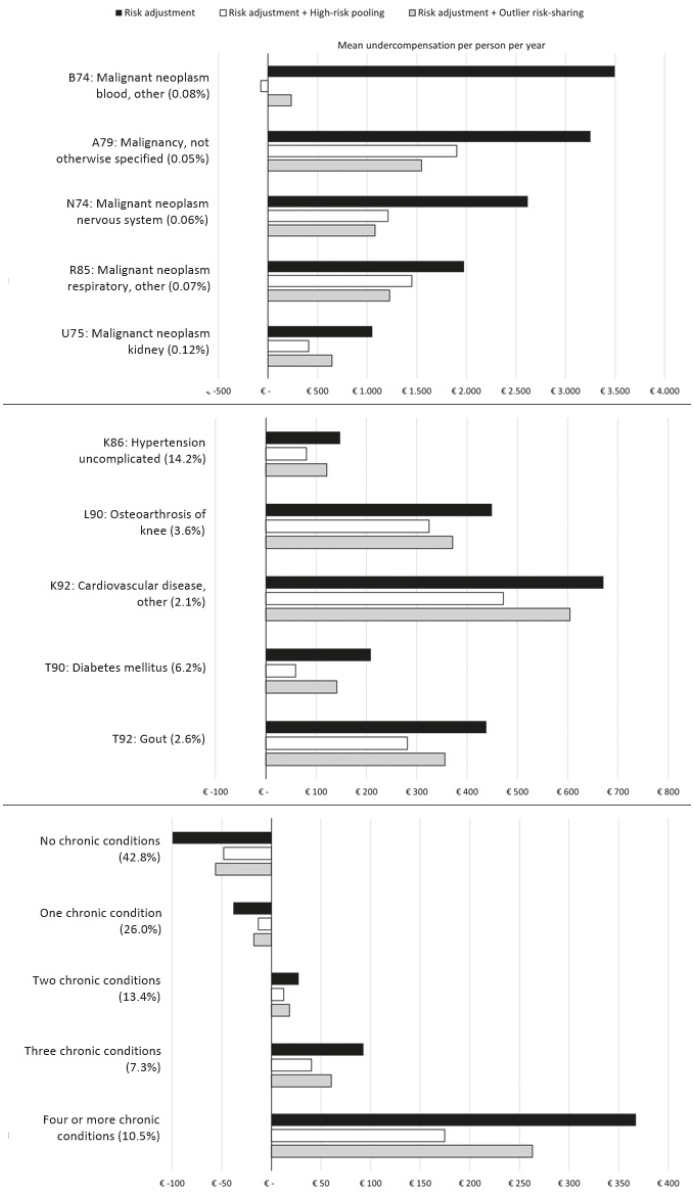
The first panel of Figure 5.2 presents the five conditions that have the largest mean undercompensation when using the RA model 2021, statistically significantly different from zero ($p < 0.01$). The black bars indicate the mean undercompensation before HRP.

Enrolling an individual with any of these conditions is clearly predictably unprofitable to insurers. With all five conditions being cancer types, insurers could be swayed towards actions that dissuade cancer patients (or specific types thereof) from enrolling. As indicated by the white bars, HRP substantially reduces the mean undercompensation for these disease groups, leaving only two means statistically significantly different from zero (only at $p < 0.05$). Notably, the largest mean undercompensation of B74 is reduced from €3,497 to an *overcompensation* of €72. Overall, of the 109 chronic illnesses in the GP data, 97 are undercompensated by RA model before HRP (of which 54 are statistically significantly undercompensated at $p < 0.05$ and 38 at $p < 0.01$). Our HRP application reduces that number to 75 (32 at $p < 0.05$ and 25 at $p < 0.01$).

The second panel in Figure 5.2 shows the effects of HRP from another angle by looking at the five diagnoses with the largest *macro undercompensations* (the mean undercompensation of a condition multiplied by its prevalence). Also, for these groups, HRP substantially reduces the mean undercompensation.

In the final panel of Figure 5.2, the mean undercompensations are shown for individuals grouped by the number of chronic illnesses reported by the GPs. Roughly 70 percent of the Dutch population has either one diagnosis or none, and, on average, those (relatively) healthy individuals are overcompensated by the RA model. As the number of chronic illnesses increases, so do the undercompensations for those groups of individuals. HRP considerably reduces the undercompensations for unhealthy individuals and the overcompensations of their healthier counterparts. This observation implies that the top one percent of individuals with the highest predicted undercompensation is responsible for a significant share of the overall undercompensation of chronically ill individuals.

Figure 5.3. Mean undercompensation per person per year for specific subgroups under the Dutch risk adjustment model 2021 before and after either high-risk pooling or outlier risk-sharing.



Note: Prevalence of the subgroup is shown between brackets. Results are based on the test sample. [Panel 1] Top 5 most severely significantly ($p < 0.01$) undercompensated diagnoses under RA model 2021 and their undercompensation before and after HRP or ORS. [Panel 2] Top 5 diagnoses with the largest macro undercompensations under RA model 2021 and their undercompensation before and after HRP or ORS. [Panel 3] Individuals grouped by the number of chronic diseases diagnosed by the GP and their undercompensation before and after HRP or ORS. The scale of the x-axis differs per panel. Prevalence of the subgroup is shown between brackets. Results are based on the test sample.

5.4.4 High-risk pooling versus outlier risk-sharing

As a final step, we compare the effects of our HRP modality with a more traditional risk-sharing modality: ORS. Using the same budget as HRP, ORS cuts off health spending at a predefined threshold. More specifically, ORS assigns those people to the pool whose spending exceeds a certain threshold. The compensation for these people equals actual spending above that threshold. Compensations are financed via a flat contribution by individuals with spending below the threshold (section 5.3.2.5). By iteratively testing, we found that with an ORS threshold of €104,325 the total ORS-budget equals the HRP-budget (i.e., roughly 500 million euros in the test sample). In total, only 0.09% of the test sample (7,905 individuals) exceeds that threshold.

For both systems we calculated the Payment System Fit (PSF), a measure that quantifies the portion of the variance in health care spending that is compensated for by a payment system (Layton et al., 2018a).[12] In the status quo model – that is, without risk-sharing – the PSF is 35.3%. For the HRP and ORS modalities, respectively, the PSF is 46.7% and 60.3%. This implies that ORS better compensates for the variance in health care spending than HRP. However, this does not necessarily mean that ORS outperforms HRP when it comes to reducing selection incentives. To indicate the impact on selection incentives, we calculated the undercompensation for subgroups of chronically ill people identified in the GP data.

For the same sets of subgroups shown in Figure 5.2, Figure 5.3 demonstrates the effects of both risk-sharing modalities on the average undercompensation. Both modalities unsurprisingly perform better than the RA model in isolation. The results between the two modalities are mixed: for some conditions ORS outperforms HRP but in general the latter leads to a larger reduction in the undercompensations on chronically ill individuals (and risk selection incentives against these groups). Of the five most undercompensated groups (first panel of Figure 5.3), three are better addressed by ORS than by HRP. For larger subgroups, HRP performs better. To facilitate a broader comparison of the two modalities, a summary measure for all 109 conditions (and the healthy individuals) is calculated for the three models: the weighted mean absolute result (WMAR).[13] For the 2021 RA model, the WMAR equals €182, while for the 2021 RA model supplemented with either HRP or ORS, the respective WMARs are €107 and €139. Therefore, when it comes to the overall reduction of selection incentives regarding the 109 conditions analyzed here, our HRP modality outperforms the more conventional ORS modality.

12 The PSF is calculated through: $1 - \frac{\sum_i (y_i - R_i)^2}{\sum_i (y_i - \hat{y}_i)^2}$. Where, in the numerator, individual-level spending y_i is subtracted by the individual-level revenue received by the insurer R_i. In a modality without explicit risk-sharing modalities outside traditional RA, this revenue is simply the RA payments received. In modalities with risk-sharing, the revenue is adjusted by the compensations received through risk-sharing and the contributions made to the modality.

13 The WMAR is calculated through: $\frac{|N_g * undercompensation_g|}{\sum_g N_g}$. Where the mean absolute undercompensation for every subgroup is multiplied by its prevalence. The sum of the results for all 109 respective conditions and the subgroup of individuals that have no condition whatsoever is then divided by the summed number of individuals in all subgroups to find a weighted deviation from zero for all subgroups combined. Since individuals can have multiple conditions at once, the subgroups in this equation are not mutually exclusive.

5.5 DISCUSSION

In this research we first studied the potential of exploiting historical administrative data on health spending and risk adjuster information to identify individuals that are predicted to be substantially unprofitable to insurers. Using prior-year spending as a predictor alone effectively indicates individuals that are substantially undercompensated by the Dutch RE model. On average, individuals in the top one percent of the spending distribution in the prior year are undercompensated by €4,918 in the current year. However, complementing prior-year spending with historical risk adjuster information in an OLS model results in a more accurate identification of 'high-risks' (€5,738 average undercompensation). Moreover, opting for a more advanced technique to exploit the information supplied – the RF algorithm – identifies a more selective group of individuals unprofitable to insurers (€6,050).

The inclusion of additional historical information in the prediction model for high-risk individuals makes a notable contribution to its strength but adds to its complexity. The model based on prior-year health spending alone is straightforward, requires no further calculations, and has predictive strength. Nevertheless, it would be shortsighted to assume competing insurers will use past spending alone in their estimations of individuals that will be undercompensated. Therefore, any information with predictive strength could be considered useful to policymakers, for instance, through potential uptake in the RA model. Later in this section, we will discuss this strategy further.

In addition to adding predictive information, we studied the effect of using a more complex method of exploiting the provided information: the RF algorithm. In comparison, the algorithm only marginally outperforms the linear regression model – introducing an important tradeoff between the slight increase in predictive performance and the compromise on transparency of the model. Therefore, if the identification of high-risk individuals as simulated in this study were to be considered for policy implementation, policymakers may opt for the more transparent and less complex model or even choose to simply use prior-year health spending. However, we do not rule out greater predictive potential for alternative, more complex machine learning models that could swing the tradeoff between performance and transparency in a different direction.

In general, the adoption of machine learning methods to identify high-risk individuals seems more viable than using it to replace the estimation of the RA model altogether. In practice, transparency and explainability of the model and the logic behind the payment coefficients of risk adjusters are valued by policymakers and competing health insurers. Therefore, traditional OLS will likely remain as the preferred estimation method. For the design of risk adjusters and classification of individuals into risk classes, however, transparency and logic are less essential. Various studies have successfully identified sub-

groups of individuals through machine learning methods and used those to improve the 'standard' estimation method for RA (Andriola et al., 2024; Zink & Rose, 2021). Our allocation of individuals to a high-risk pool fits within that branch of research. This identification of high-risk individuals through historical information could still be applied to improve RA, either through HRP or through the adoption of new risk adjusters or adaptation to existing ones, based on the applied information.

As a second part of our analysis, we simulated a form of high-risk pooling (HRP) as a supplement to RA, specifically for the individuals identified in the first part of our analysis. We find that HRP results in a substantial reduction of overall selection incentives: using 2.6% of the total budget for compensating the identified group based on residual spending, on average, reduces the undercompensation for the 109 chronic conditions by 42% (i.e., the weighted mean absolute result for the 109 chronic conditions (section 5.4.4) decreases from €182 to €107). This implies that the top one percent of individuals with the highest predicted undercompensation is responsible for a significant share of the overall undercompensation of chronically ill individuals. By better compensating this group via HRP, the incentives for insurers to select against specific chronic illnesses substantially diminish.

In a final analysis, the results of HRP are compared to those of a more conventional – and commonly used – risk-sharing modality: ORS. Using the same risk-sharing budget (2.6% of the total budget), the ORS modality is less effective at reducing selection incentives towards the chronically ill. In our analyses we simulated ORS with a flat contribution to draw a fair comparison with HRP in terms of financing. In health insurance systems with ORS, the ORS-compensations are often financed by a recalibration of the RA model on the costs that remain *after* actual spending has been top-coded at threshold T^* (affecting every individual). As an additional analysis, we simulated this more common ORS and found that HRP outperforms ORS (Figure 5A.1 in the Appendix). Moreover, the ORS with a flat contribution structurally outperforms the more traditional ORS-variant with a recalibrated RA model. The reason is that recalibrating the model on the top-coded spending substantially reduces the RA payment weights for risk adjusters that flag chronically ill individuals. So, while ORS transfers funds towards groups of chronically ill through cost-based compensations, it takes away funds from these groups through lower recalibrated RA payment weights. Opting for a financing structure that more resembles the HRP design might be a better strategy to safeguard the predictive power of the RA payment weights. This is an interesting direction for future research. Moreover, in our ORS modalities, all spending above the defined threshold is compensated, eliminating incentives for cost-containment for that spending. Alternative options could be considered to reduce this impact. For instance, a predefined percentage of coverage of spending above the threshold could be applied, as is done in some systems in Europe. The same principle holds for HRP, where a predefined share could be considered to alleviate concerns of cost-containment. However, these changes in risk-sharing alter the impact on selection

incentives by the modalities, suggesting consideration of the tradeoff between cost-containment and selection.

Irrespective of the modality chosen, risk-sharing decreases efficiency through a reduction of incentives for cost-containment. A counterargument, however, is that risk-sharing simultaneously reduces selection incentives – as shown in our study – which improves efficiency (Van Kleef et al., 2022a). The hypothesis underlying this argument is that when fewer financial gains are to be made by insurers through actions of risk selection, they will be more likely to resort to efficient contracting of care services to generate profits (Enthoven, 1988a). However, while risk-sharing in our analyses is an effective strategy to reduce risk selection incentives, undercompensations remain. Therefore, the predictable (un)profitability of individuals may still undermine the potential focus of insurers on efficient contracts.

Since RA models currently in place do not eliminate the undercompensation of subgroups of individuals vulnerable to risk selection, the uptake of historical information as used in this study could theoretically improve the RA model. However, while prior-year health spending has proven predictive strength, it is commonly regarded as unsuitable for RA due to incentives for gaming. For example, a risk adjuster that explicitly flags the group of enrollees that exceeded a spending threshold in the prior year would provide insurers with a fixed additional payment for each enrollee in that group. This can provide insurers with incentives to make sure that the individuals who approach the spending threshold indeed exceed it, for example, by providing unnecessary care.[14] This disadvantage does not apply (or applies to a lesser extent) to risk adjusters from prior years. Our analyses indicate that such information yields predictive strength, while it should be notably more difficult for insurers to game a risk adjuster based on this information. To make new gains in RA, further research should investigate the potential of historical and multi-year information on health for the configuration of risk adjusters.

An alternative solution to the issue may be loosening principles of solidarity and allowing insurers to differentiate premiums to cover for predictable losses. Different estimation methods for the RA model could be considered as well. For instance, Van Kleef et al. (2020a) and Van Kleef et al.(2022b) apply a constrained regression to reduce the undercompensation of specific subgroups of the population. Moreover, given the predictive strength of prior-year spending found in this study, it could be used in the various strategies discussed above. However, since all strategies have their own advantages and concerns, more work is required to further reduce selection incentives in health insurance markets adequately.

14 Our modality of high-risk pooling does not have this problem since the compensation from the high-risk pool can only reduce the loss of an insurer for an enrollee but will never increase the profit made on an enrollee.

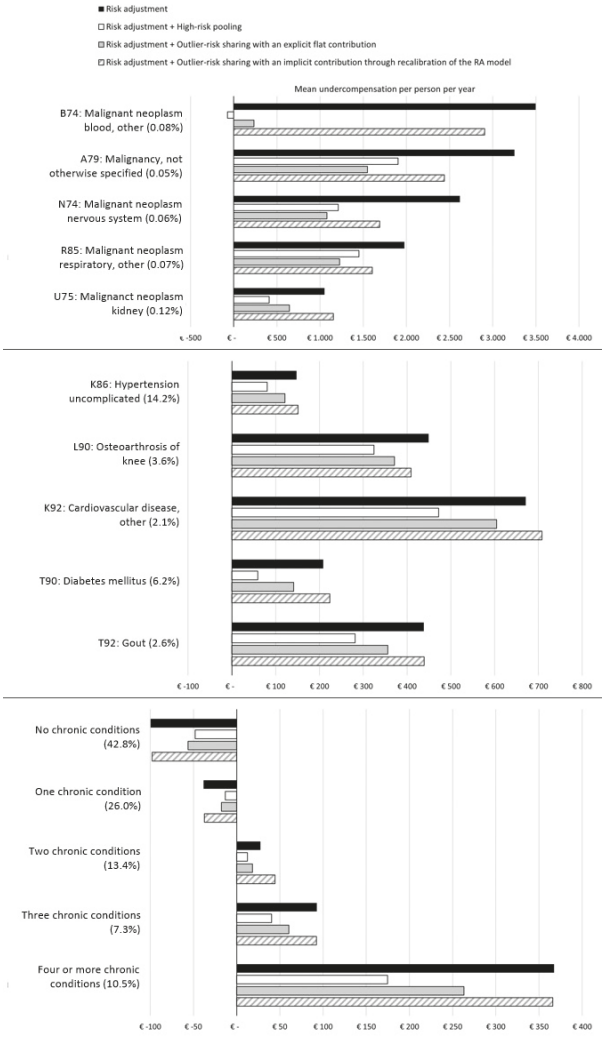
Within this research, the information supplied by GPs played a crucial role. To evaluate payment systems in the context of health insurance, the quantification of selection incentives towards relevant subgroups of the population is a key element. The GP information on the health status of individuals facilitated the grouping of individuals that are potentially vulnerable to risk selection by health insurers. Although general statistical measures such as PSF provide valuable information for evaluation – in the case of PSF with respect to the fit of a payment system in relation to the variation in health spending –, they do not directly indicate selection incentives. The undercompensations of the specific subgroups derived from the GP data form a direct example of how selection incentives remain within the market and how payment systems can be compared on an important, tangible scale.

This paper demonstrates the potential effectiveness of HRP as a supplement to RA to reduce selection incentives. However, its strength could be further increased through refinement of the instrument. For instance, with more (extensive) data or different estimation methods, the identification of high-risk individuals could potentially be improved. For instance, residual spending in prior years is likely to be correlated to the residual spending in future years. This is an interesting subject for further research to address consistency in undercompensations and improve RA altogether but falls outside the scope of this study. Clearly, more information than that used in this study will have (more) predictive strength with respect to residual spending. It remains pivotal for the functioning of RA and the health insurance market to explore suitable streams of predictive, detailed information to reduce selection incentives. Also, increasing the size of the pool of individuals qualifying for HRP could work to further decrease overall undercompensations but would require a larger budget. Alternatively, the financing of the compensation could be overhauled. Financing the payments through flat contributions is an arbitrary choice and could be accommodated differently, for example, by taking away funds from the group of enrollees with the largest *overcompensations* or by imposing a (relative) deduction from RA payments (McGuire et al., 2021a). The implication of such modifications for the outcomes of HRP is an interesting empirical topic for further research.

Nevertheless, the findings of this study show that HRP is a promising instrument for the design of health plan payment systems. Directing supplementary compensation towards a selective group of identified high-risk individuals substantially reduces the selection incentives on the insurance market. While risk-sharing has its inherent caveats, HRP has the potential to improve the functioning of health insurance systems. Especially considering the understanding that competing insurers can identify high-risk individuals themselves, HRP could make an effective strategy to reduce incentives for risk selection. Last, the findings of this study are not exclusively valuable for the Netherlands. Provided RA models elsewhere similarly undercompensate chronically ill individuals, the implications of our research can be insightful to policymakers in Germany, Switzerland, the US, and other countries employing RA.

5.6 APPENDIX

Figure 5A.1. Mean undercompensation per person per year for specific subgroups under the Dutch risk adjustment model 2021 before and after either high-risk pooling, outlier-risk-sharing or traditional outlier-risk-sharing with a recalibration of the risk adjustment model.



Note: Prevalence of the subgroup is shown between brackets. Results are based on the test sample. [Panel 1] Top 5 most severely significantly ($p < 0{,}01$) undercompensated diagnoses under RA model 2021 and their undercompensation before and after HRP or two variants of ORS. [Panel 2] Top 5 diagnoses with the largest macro undercompensations under RA model 2021 and their undercompensation before and after HRP or two variants of ORS. [Panel 3] Individuals grouped by the number of chronic diseases diagnosed by the GP and their undercompensation before and after HRP or two variants of ORS. The scale of the x-axis differs per panel. Prevalence of the subgroup is shown between brackets. Results are based on the test sample.

Table 5A.1. An overview of the dummy-variables computed for health spending in year *t-1*, used for Ordinary Least Squares regression and Random Forest algorithm.

Health spending in year *t-1*	Dummy-variable
Above €209,268	99.99-100th fraction
Between €160,841 and €209,268	99.98-99.99th fraction
Between €135,971 and €160,841	99.97-99.98th fraction
Between €122,174 and €135,971	99.96-99.97th fraction
Between €111,913 and €122,174	99.95-99.96th fraction
Between €104,119 and €111,913	99.94-99.95th fraction
Between €98,572 and €104,119	99.93-99.94th fraction
Between €94,050 and €98,572	99.92-99.93th fraction
Between €90,229 and €94,050	99.91-99.92th fraction
Between €86,962 and €90,229	99.90-99.91th fraction
Between €60,221 and €86,962	99.75-99.90th fraction
Between €43,645 and €60,221	99.5-99.75th fraction
Between €35,339 and €43,645	99.25-99.5th fraction
Between €29,922 and €35,339	99-99.25th fraction
Between €23,062 and €29,922	98.5-99th fraction
Between €18,811 and €23,062	98-98.5th fraction
Between €13,930 and €18,811	97-98th fraction
Between €11,088 and €13,930	96-97th fraction
Between €9,125 and €11,088	95-96th fraction
Between €7,624 and €9,125	94-95th fraction
Between €6,505 and €7,624	93-94th fraction
Between €5,657 and €6,505	92-93th fraction
Between €4,999 and €5,657	91-92th fraction
Between €4,474 and €4,999	90-91th fraction
Between €1,446 and €4,474	75-90th fraction
Between €414 and €1,446	50-75th fraction
Below €414	First 50%

Table 5A.2. The mean residual spending for the top five percentiles of predicted residual spending based on various combinations of hyperparameters of the random forest algorithm

Number of trees	Minimal node size	96th percentile	97th percentile	98th percentile	99th percentile	100th percentile
200	25	€639	€713	€777	€1,410	€5,786
200	50	€559	€726	€905	€1,386	€5,879
200	75	€609	€703	€832	€1,442	€5,943
200	100	€576	€714	€826	€1,452	€5,948
200	150	€607	€652	€882	€1,435	€6,050
200	200	€626	€679	€836	€1,474	€6,035
200	300	€627	€669	€913	€1,481	€6,035
200	500	€566	€701	€810	€1,509	€6,035

Table 5A.3. Descriptive statistics of the subgroups separated by high-risk pooling design

	Test sample	Bottom 99%	Top 1%
Number of insured	8,628,147	8,535,293	92,854
Sum of insured years	8,484,607	8,399,761	84,846
Mean spending year *t*	€2,407	€2,119	€30,950
Mean predicted spending year *t*	€2,407	€2,180	€24,900
Mean residual spending year *t*	€0	€ - 61	€6,050
Frequencies			
Women	50.5%	50.6%	46.3%
Ages 1 to 17	18.8%	19.8%	7.4%
Ages 18 to 34	20.8%	21.0%	3.8%
Ages 35 to 44	11.9%	12.0%	3.2%
Ages 45 to 54	14.8%	14.9%	7.3%
Ages 55 to 64	13.6%	13.5%	15.4%
Ages 65 to 74	11.1%	11.0%	22.2%
Ages 75 and older	8.1%	7.8%	40.7%
PCGs > 0	16.8%	16.2%	74.0%
DCGs > 0	11.4%	10.7%	82.4%
DMECGs > 0	4.5%	4.1%	42.1%
PDGs > 0	2.0%	1.8%	15.1%
Multiple-year high-costs	6.1%	5.5%	63.5%
Prior spending on home care	2.4%	2.0%	39.8%
At least in one of the above morbidity classes	25.2%	24.5%	97.2%
Mean spending year *t-1*	€2,183	€1,697	€46,772

Table 5A.4. All 109 chronic conditions from the GP data with their respective mean undercompensation under (1) the Dutch risk adjustment scheme of 2021, (2) the adjustment scheme supplemented with high-risk pooling, and (3) the adjustment scheme supplemented with outlier-risk-sharing.

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean undercompensation per person per year (€)	Mean undercompensation per person per year with high-risk pooling(€)	Mean undercompensation per person per year with outlier-risk-sharing(€)
General and Unspecified						
A28	Limited function/disability, not otherwise specified	0.13	6,924	255	145	233
A79	Malignancy, not otherwise specified	0.053	17,886	** 3,252	* 1,903	* 1,549
A90	Congenital anomaly, not otherwise specified/multiple	0.19	4,994	185	- 135	- 119
Blood, Blood Forming Organs and Immune Mechanism						
B28	Limited function/disability	0.023	6,251	- 856	* - 893	- 795
B72	Hodgkin's disease/lymphoma	0.20	9,451	361	- 294	29
B73	Leukaemia	0.14	14,838	- 556	** - 2,914	- 2,175
B74	Malignant neoplasm blood, other	0.082	27,734	** 3,497	- 72	234
B78	Hereditary haemolytic anaemia	0.30	3,033	157	38	105
B79	Congenital anomaly blood/lymph, other	0.052	7,243	629	- 30	- 108
B83	Purpura/coagulation defect	0.71	6,164	106	** - 386	** - 335
B90	HIV-infection/aids	0.12	13,962	394	111	171
Digestive system						
D28	Limited function/disability	0.053	9,730	124	* - 289	- 365
D74	Malignant neoplasm stomach	0.052	10,850	272	* - 328	- 126
D75	Malignant neoplasm colon/rectum	0.73	9,492	* 299	- 16	116
D76	Malignant neoplasm pancreas	0.029	16,172	156	- 375	- 279
D77	Malignant neoplasm digestive, other/not otherwise specified	0.18	11,008	** 983	338	* 662
D81	Congenital anomaly digestive system	0.37	2,165	* 186	143	122
D92	Diverticular disease intestines	1.87	6,233	** 256	** 180	** 233
D94	Chronic enteritis/ulcerative colitis	0.87	6,588	*173	53	102
D97	Cirrhosis/Liver disease, not otherwise specified	0.79	6,032	** 557	** 305	** 370

Table 5A.4. All 109 chronic conditions from the GP data with their respective mean undercompensation under (1) the Dutch risk adjustment scheme of 2021, (2) the adjustment scheme supplemented with high-risk pooling, and (3) the adjustment scheme supplemented with outlier-risk-sharing. (continued)

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean undercompensation per person per year (€)	Mean undercompensation per person per year with high-risk pooling(€)	Mean undercompensation per person per year with outlier-risk-sharing(€)
Eye						
F28	Limited function/disability	0.23	4,627	206	182	214
F81	Congenital anomaly eye, other	0.22	2,786	192	198	113
F83	Retinopathy	0.70	9,507	** 331	21	151
F84	Macular degeneration	0.68	9,054	188	73	152
F91	Refractive error	4.39	2,556	** 82	** 74	** 93
F93	Glaucoma	1.69	5,324	105	21	61
F94	Blindness	0.20	5,770	93	- 17	98
Ear						
H28	Limited function/disability	0.12	4,912	* 533	* 531	** 564
H80	Congenital anomaly of ear	0.14	2,590	* 353	* 293	** 334
H83	Otosclerosis	0.083	3,521	52	29	88
H84	Presbycusis	1.74	7,177	* 145	17	100
H85	Acoustic trauma	0.34	4,265	148	44	117
H86	Deafness	2.19	5,060	** 274	** 171	** 223
Cardiovascular						
K28	Limited function/disability	0.13	5,681	* 539	* 389	* 509
K73	Congenital anomaly cardiovascular	0.40	3,697	112	111	81
K74	Angina pectoris	2.37	7,783	* 144	3	98
K76	Other and chronic ischaemic heart disease	1.11	8,183	** 490	** 294	** 430
K77	Heart failure	1.09	13,660	** 441	82	183
K82	Pulmonary heart disease	0.051	17,962	* 1,560	- 168	- 168
K86	Hypertension uncomplicated	14.16	5,166	** 148	** 81	** 121
K87	Hypertension complicated	1.90	8,168	** 516	** 351	** 378
K90	Stroke/cerebrovascular accident	1.83	8,849	** 458	** 209	** 371
K91	Atherosclerosis	1.33	6,469	** 328	** 223	** 288
K92	Other arterial obstruction / peripheral vascular disease	2.09	7,658	** 671	** 472	** 605

Table 5A.4. All 109 chronic conditions from the GP data with their respective mean undercompensation under (1) the Dutch risk adjustment scheme of 2021, (2) the adjustment scheme supplemented with high-risk pooling, and (3) the adjustment scheme supplemented with outlier-risk-sharing. (continued)

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean undercompensation per person per year (€)	Mean undercompensation per person per year with high-risk pooling(€)	Mean undercompensation per person per year with outlier-risk-sharing(€)
Musculoskeletal						
L28	Limited function/disability	0.38	7,286	** 594	** 386	** 472
L82	Congenital anomaly musculoskeletal	1.24	2,194	81	* 93	45
L84	Osteoarthritis of spine (any region)	1.56	6,209	** 210	106	** 176
L85	Acquired deformity of spine	0.93	3,389	*138	* 116	** 139
L88	Rheumatoid arthritis and allied conditions	1.39	7,176	82	- 65	5
L89	Osteoarthrosis of hip	2.44	6,966	** 258	** 145	** 203
L90	Osteoarthrosis of knee	3.61	6,762	** 450	** 325	** 371
L91	Osteoarthrosis, other	3.21	5,518	** 172	** 120	** 143
L95	Osteoporosis	2.65	6,982	** 260	** 133	** 180
L98	Acquired deformity of limb	4.29	3,270	** 140	** 120	** 121
Neurological						
N28	Limited function/disability	0.047	6,299	454	269	506
N70	Poliomyelitis/other enterovirus	0.045	6,012	* - 723	* - 676	* - 674
N74	Malignant neoplasm nervous system	0.061	12,519	** 2,619	1,212	* 1,081
N85	Congenital anomaly neurological	0.12	8,250	438	- 41	- 18
N86	Multiple sclerosis	0.19	12,086	246	23	220
N87	Parkinsonism/paralysis agitans	0.28	12,519	355	- 88	317
N88	Epilepsy, all types	1.09	5,562	* 199	6	20
Psychological						
P28	Limited function/disability	0.10	4,123	217	178	189
P70	Dementia	0.61	7,390	** - 1,208	** - 1,329	** - 1,217
P72	Schizophrenia, all types	0.28	3,723	** - 259	* - 242	* - 223
P80	Personality disorder	1.14	3,379	** 195	** 186	** 161
P85	Mental retardation	0.59	3,287	- 45	- 64	- 41

Table 5A.4. All 109 chronic conditions from the GP data with their respective mean undercompensation under (1) the Dutch risk adjustment scheme of 2021, (2) the adjustment scheme supplemented with high-risk pooling, and (3) the adjustment scheme supplemented with outlier-risk-sharing. (continued)

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean undercompensation per person per year (€)	Mean undercompensation per person per year with high-risk pooling(€)	Mean undercompensation per person per year with outlier-risk-sharing(€)
Respiratory						
R28	Limited function/disability	0.079	8,914	571	140	- 339
R84	Malignant neoplasm trachea/ bronchus/lung	0.23	17,566	*800	* - 713	188
R85	Malignant neoplasm respiratory, other	0.069	11,747	** 1,976	* 1,451	** 1,227
R89	Congenital anomaly respiratory	0.036	8,058	* 1,829	- 130	124
R91	Chronic bronchitis/ bronchiectasis	0.76	6,008	190	59	133
R95	Emphysema/chronic obstructive pulmonary disease	2.59	8,471	** 385	** 147	** 274
R96	Asthma	9.63	3,034	** 60	** 49	** 68
Skin						
S28	Limited function/disability	0.061	4,146	167	- 46	159
S77	Malignant neoplasm of skin	3.60	5,770	** 184	71	** 112
S81	Haemangioma/lymphangioma	0.93	2,698	97	82	80
S83	Congenital skin anomaly, other	0.30	2,689	10	4	22
S87	Atopic dermatitis/eczema	10.65	2,173	** 41	** 59	** 58
S91	Psoriasis	2.59	4,269	*87	55	** 92
Endocrine/Metabolic and Nutritional						
T28	Limited function/disability	0.011	8,097	2,180	883	937
T71	Malignant neoplasm thyroid	0.063	6,055	219	- 20	12
T78	Thyroglossal duct cysts	0.10	2,970	114	58	4
T80	Congenital anomaly endocrine/ metabolic	0.11	8,081	* 957	190	* - 490
T81	Goitre, thyroid nodule including thyrotoxicosis	0.53	4,704	99	103	123
T86	Hypothyroidism/myxoedema	2.72	4,756	27	5	0
T90	Diabetes mellitus	6.20	7,298	** 209	59	** 141
T92	Gout	2.63	6,550	** 438	** 282	** 356
T93	Lipid metabolism disorder	7.95	4,628	** 115	** 77	** 108

Table 5A.4. All 109 chronic conditions from the GP data with their respective mean undercompensation under (1) the Dutch risk adjustment scheme of 2021, (2) the adjustment scheme supplemented with high-risk pooling, and (3) the adjustment scheme supplemented with outlier-risk-sharing. (continued)

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean undercompensation per person per year (€)	Mean undercompensation per person per year with high-risk pooling(€)	Mean undercompensation per person per year with outlier-risk-sharing(€)
Urological						
U28	Limited function/disability urinary	0.069	16,531	716	- 532	- 6
U75	Malignant neoplasm of kidney	0.12	12,183	** 1,053	411	* 644
U76	Malignant neoplasm of bladder	0.28	10,021	*358	- 36	178
U77	Malignant neoplasm urinary, other	0.018	10,729	360	49	301
U85	Congenital anomaly urinary tract	0.18	4,802	** 491	** 434	** 435
U88	Glomerulonephritis/nephrosis	0.15	8,472	** 1,006	** 655	** 655
Pregnancy, Childbearing, Family Planning						
W28	Limited function/disability	0.12	2,214	50	100	105
W72	Malignant neoplasm coexisting with pregnancy	0.0012	1,827	- 183	- 121	- 122
W76	Congenital anomaly complicating pregnancy	0.015	2,346	40	47	31
Female Genital						
X28	Limited function/disability	0.0076	3,151	- 142	- 81	- 82
X75	Malignant neoplasm cervix	0.25	4,265	181	150	218
X76	Malignant neoplasm breast female	1.36	6,795	** 249	111	** 213
X77	Malignant neoplasm female genital, other	0.26	7,315	326	107	307
X83	Congenital anomaly genital female	0.062	2,240	- 2	59	59
X88	Chronic cystic disease breast female	1.52	2,870	** 164	** 175	** 167

Table 5A.4. All 109 chronic conditions from the GP data with their respective mean undercompensation under (1) the Dutch risk adjustment scheme of 2021, (2) the adjustment scheme supplemented with high-risk pooling, and (3) the adjustment scheme supplemented with outlier-risk-sharing. (continued)

Code	Chronic condition	Prevalence (%)	Mean spending per person per year (€)	Mean undercompensation per person per year (€)	Mean undercompensation per person per year with high-risk pooling(€)	Mean undercompensation per person per year with outlier-risk-sharing(€)
Male Genital						
Y28	Limited function/disability	0.11	4,972	16	- 12	55
Y77	Malignant neoplasm prostate	0.61	9,177	** 432	144	** 329
Y78	Malignant neoplasm male genital, other	0.11	3,905	- 18	- 69	- 5
Y82	Hypospadias	0.10	1,647	- 57	- 18	4
Y84	Congenital male genital anomaly, other	0.056	1,510	- 52	9	8
Social Problems						
Z28	Limited function/disability	0.34	3,354	2	- 84	- 56

Note: Results are based on the rebalanced sample from Dutch GPs (N = 1.5m). As explained in footnote 8, all individuals from that sample were allocated to the test sample. Conditions for which the undercompensation in the three scenarios is statistically significantly different from zero are marked with an asterisk. ** = statistically significantly different from zero with $p < 0.01$. * = statistically significantly different from zero with $0.01 < p < 0.05$.

6

MITIGATING INCENTIVES FOR RISK SELECTION BY RELAXING PREMIUM-RATE RESTRICTIONS IN THE PRESENCE OF RISK EQUALIZATION

Based on: Oskam, M., Van Kleef, R. C., & Van de Ven, W. P. (2026). To What Extent Does State-of-the-Art Risk Equalization Create Cross-Subsidies on a Competitive Health Insurance Market? *Submitted for publication.*

Abstract

Regulators of social health insurance markets typically implement premium-rate restrictions (e.g., community-rating) to achieve implicit cross-subsidization from low-risk to high-risk individuals. A drawback of premium-rate restrictions is that they create incentives for insurers to engage in risk selection. Many countries use risk equalization (RE) to counteract selection incentives. An interesting question is whether regulators need premium-rate restrictions in the presence of state-of-the-art RE given that RE by itself generates (explicit) cross-subsidies from low- to high-risk individuals. This paper empirically examines to what extent state-of-the-art RE creates cross-subsidization in the absence of premium-rate restrictions. Our analysis is focused on the Dutch basic health insurance and the Dutch RE model for somatic care of 2022. We find that this RE model achieves 81% of the cross-subsidies that would be achieved by implementing community-rating across the market, subject to several restrictive assumptions. We discuss the pros and cons of premium-rate restrictions as well as other potential instruments to further increase the cross-subsidies. Our results may help regulators evaluate whether, in the case of state-of-the-art RE, the marginal benefits of premium-rate restrictions (the additional cross-subsidization) outweigh the marginal costs (the potential effects of regulation-induced selection).

6.1 INTRODUCTION

Unregulated, competitive insurance markets typically tend towards equivalence between the premium for coverage and the expected cost of insurance contracts (Van de Ven & Ellis, 2000a). Consequently, premiums will be higher for contracts with high expected costs than for contracts with low expected costs. For car or fire insurance, such premium differentiation between risk types is commonly accepted and applied. For health insurance, however, such differentiation is not as socially accepted. Applying risk-rated premiums in health insurance could lead to substantial premiums for chronically ill and elderly individuals, threatening the affordability of coverage.

To enhance affordability of coverage, regulators of *social* health insurance markets typically implement premium-rate restrictions (often as community-rating per insurance product or across the market) combined with an open enrollment requirement for a standardized benefits package (i.e., a prohibition for insurers to reject applicants for health insurance) (Van Kleef et al., 2018b). Such premium-rate restrictions create *implicit* cross-subsidies between the various risk types on the market: low-risks (for whom the community-rated premium exceeds the expected costs) implicitly subsidize the premium of high-risks (for whom the community-rated premium falls below the expected costs). A well-known disadvantage of these implicit cross-subsidies is that they lead to predictable profits for insurers on low-risk individuals and predictable losses on high-risk individuals.

Without additional policy measures, predictable profits and losses between risk types can result in risk selection: actions to exploit the unpriced risk heterogeneity and break pooling arrangements (Newhouse, 1996). Examples of selection actions include marketing strategies or distortions of provider networks, where high-quality providers of specific types of care are not contracted by the insurer to deter the enrollment of individuals in need of such services (Glazer & McGuire, 2000; Van de Ven et al., 2015; Shepard, 2022; Stolper et al., 2022). Such regulation-induced 'risk selection' could hamper (investment into) efficiency and quality of care, threatening the functioning of the health insurance market as well as the functioning of the broader health care system (Newhouse, 1996; Einav et al., 2024). This is because risk selection may also induce price distortions and inefficient sorting of individuals across health plans: when high-risk individuals gravitate towards specific plans, premium variations might reflect selection effects rather than the quality of the insurance plan, distorting the sorting of consumers across insurance plans and distorting the level playing field among insurers (Glazer & McGuire, 2000; Van de Ven & Ellis, 2000a). Moreover, efficiency of production is hindered if risk selection proves to be a more effective strategy for cost-control than aligning with consumer preferences (Van de Ven & Ellis, 2000a; Glazer & McGuire, 2000).

To mitigate selection problems, regulators often supplement the *implicit* cross-subsidies of premium-rate restrictions with *explicit* cross-subsidies through risk equalization (RE) (Ellis et al., 2018). This instrument is designed to compensate insurers for the predictable variation in health spending of consumers. For example, insurers receive a risk-adjusted subsidy through RE that is higher for elderly and chronically ill individuals than for the young and healthy. These risk-adjusted subsidies, typically based on a prediction of health care spending, reduce the predictable profits and losses that result from premium regulations. Although the sophisticated RE models currently in place – for instance, in the Netherlands, Germany, and the Marketplaces in the United States – are significantly more effective at predicting individual health spending than the initial models of the 1990s, predictable profits and losses do remain, and so do selection incentives (Van Kleef et al., 2013b; Van Kleef et al., 2019; McGuire et al., 2021a; Oskam et al., 2023).

Alternatively, to avoid regulated-induced selection incentives, premium-rate restrictions could theoretically be removed whilst maintaining the risk-adjusted subsidies from RE. In that scenario, the risk-rated premiums that would reflect the expected cost of an insurance contract should be impacted by the presence of the subsidies. Assuming the risk-adjusted subsidies are higher for high-risk consumers than for low-risks, the differences between the risk-rated premiums for these risk types should shrink. Therefore, an alternative approach for regulators of health insurance systems could be to refrain from premium-rate restrictions and consider the RE model as the standalone instrument for creating cross-subsidization between risk types.

In an exploratory analysis performed 25 years ago, Van de Ven et al. (2000b) simulated the described scenario, assuming a premium model based on age, sex, prior hospital admissions, and prior health spending. The authors simulated a subsidy model to create risk-adjusted subsidies to compensate for the variation in premiums and achieve cross-subsidization. Within such an analysis, the success of the risk-adjusted subsidies to reduce premium variation depends on the combination of two factors: to what extent premium-variation is an accurate reflection of the risk of health spending of the consumer, and to what extent the risk-adjusted subsidies do so. Through sufficient subsidization of premium-variation, the need for premium-rate restrictions is eliminated, avoiding their adverse effect of regulation-induced selection. Both the premium model and the subsidy model used in the study by Van de Ven et al. (2000b) can be regarded as relatively limited by modern standards but present an interesting starting point for new research. In the years that have passed since the publication of the paper, a wealth of additional risk factors and health-related information has become available, which can be used by insurers for risk classification in their premium models. Therefore, in this paper we exploit extensive information to simulate a sophisticated premium model to find how a realistic subsidy model, as used in practice through RE, performs to achieve cross-subsidization. Through that strategy we seek to answer the following question: To what extent does a state-of-

the-art RE model create the cross-subsidies that would be achieved by community-rating across the market?

The relevance of our research question is that, although we know that RE does not compensate for all variation in individual health spending (e.g., Oskam et al., 2024), we have yet to quantify to what extent state-of-the-art RE actually does generate cross-subsidies. Because these cross-subsidies will be 'imperfect', we discuss the pros and cons of premium-rate restrictions, (imperfect) RE and other policy instruments to realize cross-subsidies. Our results raise the question of whether, in the presence of state-of-the-art RE, the marginal benefits of premium-rate restrictions (the additional cross-subsidization on top of the cross-subsidization achieved by RE) outweigh the marginal costs (the potential effects of regulation-induced selection). This research thus seeks to challenge the need for premium-rate restrictions as a principal regulatory intervention in health insurance markets with state-of-the-art RE.

In the next section, we develop a measure to quantify the level of cross-subsidization achieved by RE. After that, we describe the setting, data, and methods as used for the simulation analysis. Finally, we present the results of our analysis and discuss the findings and policy implications of our research.

6.2 CROSS-SUBSIDIZATION ACHIEVED BY RISK EQUALIZATION

To quantify the extent to which risk-adjusted subsidies achieve cross-subsidization, we develop a benchmarking measure in this section. For simplicity we make the following assumptions:[1]

1. insurance contracts cover the annual health care spending of individuals (not groups or families);
2. there is a standardized benefits package, a strong insurance mandate, and an open enrollment policy (but no premium regulation); [2]
3. competition incentivizes insurers to charge actuarially fair premiums that exclusively cover the expected costs of an insurance contract and do not include any loading fees;
4. insurers have information about *all* ex-ante risk characteristics of all individuals and determine the expected health care spending for each individual on the basis of these characteristics;
5. both the insurers and the RE-regulator use ordinary least squares (OLS) to predict individual health expenses.

1 In the discussion section we explain the impact of these assumptions on the findings of our research.
2 Through this assumption, we reduce the central problem of this research to one of affordability of coverage.

Within this section, we contemplate how risk-rated premiums are shaped in the absence of risk-adjusted subsidies and how these subsidies alter the premiums.

6.2.1 Risk-rated premiums in the absence of risk-adjusted subsidies

In the absence of risk-adjusted subsidies (and under the assumptions made), the risk-rated premium p for individual i will simply equal the expected health care spending for i, as predicted by the insurer[3]:

$$p_i = \hat{y}_i \tag{6.1}$$

In Figure 6.1, a theoretical distribution of these risk-rated premiums is shown. The horizontal axis in the figure denotes the premium level, and the dashed vertical line ($\bar{p}$) indicates the average premium (equal to the mean expected health care spending in the population, $\hat{\bar{y}}$). The vertical axis indicates the frequency of individuals that are charged the premium level shown on the horizontal axis. Individuals with below-average expected health care spending (i.e., those left of $\bar{p}$) are charged a below-average premium, while individuals with above-average expected health care spending (i.e., those right of $\bar{p}$) are charged an above-average premium. In this scenario, risk-rated premiums for high-risk individuals will (far) exceed those of low-risk individuals.

Figure 6.1. Distribution of risk-rated premiums in the absence of subsidies

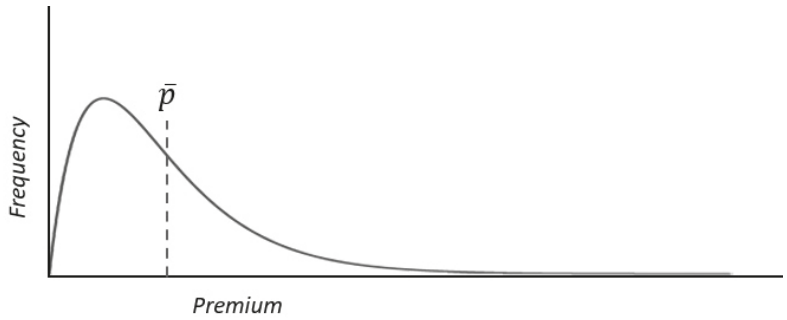


Figure 6.1 illustrates a scenario without any cross-subsidization (remember that our assumptions imply that all risk characteristics are reflected in premiums). An interesting consideration is what the premiums would look like in the presence of complete cross-subsidization. The answer is simple: complete cross-subsidization implies that the

3 In reality, competing insurers will have their own information on their respective portfolio of insured individuals to estimate expected health care spending per individual. Moreover, the insurers will likely apply varying methods to exploit the information available to them, leading to different estimations per insurer for the same applicant for insurance coverage. For this study, however, we assume the use of a single method for the available information.

premium equals $\overline{y}$ for all individuals.[4] Individuals with low expected health care spending will indirectly subsidize those with high expected health care spending through the risk-adjusted subsidies. By knowing the premiums in the two extreme scenarios of 'complete risk-rating' and 'no risk-rating at all because of complete cross-subsidization', we can calculate the sum of cross-subsidies (SCS) in a situation of complete cross-subsidization (i.e., community-rating across the market) as:

$$SCS_{complete} = \sum_i^n \left| \hat{y}_i - \overline{\hat{y}} \right| \qquad (6.2)$$

Here, we take the sum of the absolute differences between individual-level expected health care spending and mean expected health care spending across the market. This summation represents the overall redistribution of funds required to achieve complete cross-subsidization between risk types.[5]

6.2.2 Risk-rated premiums in the presence of risk-adjusted subsidies

Next, we illustrate an alternative scenario with explicit risk-adjusted subsidies to insurers. We assume that these subsidies can be either positive or negative and that they sum to zero. This implies insurers contribute to the subsidy fund for low-risk individuals while they receive from the fund for high-risk individuals. More specifically, we assume that the subsidy s, received by an insurer for individual i, equals:

$$s_i = REP_i - \overline{REP} \qquad (6.3)$$

in which REP_i refers to the prediction of individual-level health spending by the risk equalization (RE) model and $\overline{REP}$ refers to the average RE prediction (which – under the assumptions made – implies that $\overline{REP}$ equals $\overline{\hat{y}}$). In this scenario, we assume the subsidies are known to the insurers and thus taken into consideration when determining their premiums. This implies that, from the insurers' perspective, these subsidies change the expected costs of insurance contracts. The expected costs no longer equal expected health care spending $\hat{y}$ but expected health care spending minus subsidy s. Under the assumptions made, the premium p for i will now be:

$$p_{i,sub} = \hat{y}_i - s_i \qquad (6.4)$$

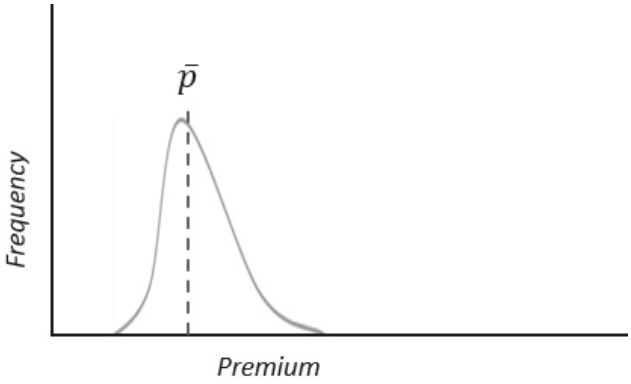
Figure 6.2 shows the theoretical distribution of risk-rated premiums after incomplete risk-adjusted subsidies have been incorporated. Compared to Figure 6.1, the distribution of premiums is strongly converged towards the average. Nevertheless, because of the

4 Because of the ordinary least squares method $\overline{y}$ equals $\overline{\hat{y}}$.

5 Since the positive subsidies will be financed by the negative subsidies in a scenario of cross-subsidization, the actual redistribution of funds should be divided by two. For this example, the absolute differences as used in Equation 6.2 are sufficient to construct the benchmark.

zero-sum principle for the subsidies, the average premium remains the same. The extent to which risk-adjusted subsidies reduce premium variation depends on the richness of the RE model used to predict health care spending.

Figure 6.2. Distribution of risk-rated premiums in the presence of (incomplete) risk-adjusted subsidies



6.2.3 A measure of realized cross-subsidies

In Figure 6.2 the sum of the absolute differences between the individual-level premiums and the mean premium in the market represents the gap between the actual level of cross-subsidization and complete cross-subsidization. This means that the cross-subsidies realized by RE as a proportion of the (complete) cross-subsidies (RPCS) can be calculated as:

$$RPCS = 1 - \frac{\sum_i^n \left| p_{i,sub} - \overline{p}_{sub} \right|}{\sum_i^n \left| \hat{y}_i - \overline{\hat{y}} \right|} \tag{6.5}$$

in which $\sum_i^n \left| \hat{y}_i - \overline{\hat{y}} \right|$ corresponds to the scenario of complete cross-subsidization (see Equation 6.2) and $\sum_i^n \left| p_{i,sub} - \overline{p}_{sub} \right|$ corresponds to the level of incomplete cross-subsidization. A value of 1 indicates full cross-subsidization. The combination of Equations 6.2 through 6.5 implies that this situation will occur if the subsidy formula perfectly mimics the formula that insurers use for calculating expected spending in premium-setting (i.e., $REP_i = \hat{y}_i$ and $\overline{REP} = \overline{\hat{y}}$). A value of 0 from Equation 6.5 indicates no cross-subsidization, which will occur in the absence of risk-adjusted subsidies.

This proposed procedure shares similarities with the work by Van Barneveld et al. (2000), Lamers (2001), and Shen and Ellis (2002), who have estimated the potential profit for insurers through risk selection. The authors simulated various RE models and used additional insurer information to identify individual-level expected profits and losses in the presence of subsidies from RE. These predictable profits and losses correspond to the

cross-subsidization that is *not* achieved. Therefore, the RPCS presented in our research can additionally serve as a measure of the extent to which a RE model reduces risk selection incentives on markets with community-rating across the market.

6.3 SETTING, DATA, AND METHODS

In this section, we describe the setting, data, and methods used within our simulation analysis. To answer our research question, we use data from the Netherlands.

6.3.1 Setting of our simulation analysis

The setting for our research is the Dutch basic health insurance scheme. This scheme is based on regulated competition and includes an insurance mandate, a standardized benefits package, premium-rate restrictions per insurance product, an open enrollment requirement, and, crucially, a sophisticated RE model that compensates insurers for predictable profits and losses that result from the premium regulations.

The RE model in place is considered state-of-the-art and includes a wide range of risk adjusters as predictors of individual-level health spending. These risk adjusters range from variables based on demographic information, such as the age and sex of individuals, to variables based on health spending from prior years, prescribed drugs in the previous year, and hospital treatments in the previous year, among others (for more information, see: Van Kleef et al. (2018a)). The latter three risk adjusters are referred to as 'morbidity adjusters', designed to identify and compensate people with chronic diseases. All risk adjusters include multiple risk classes, resulting in a total number of over 200 risk classes. Table 6.1 provides a description of all risk adjusters in the Dutch RE model for 2022, which is used as a baseline for this research. Most of these risk adjusters are 'prospective', meaning that information from preceding years is used to predict health spending in the following years. Moreover, it means that the information is known before the start of the year to which the RE payment applies. The payment weights of the 2022 RE model are estimated on the population enrolled in the basic health insurance in 2019 (year *t*), meaning that the information that underlies the morbidity adjusters stems from 2018 (year *t-1*) and before.

Table 6.1. Overview of the risk adjusters included in the Dutch risk equalization model for 2022

Risk adjuster	Operationalization
Age interacted with sex	21 classes for men and 21 classes for women, based on the age in year *t*. Most classes are 5-year cohorts, with children and individuals aged 90 years or older in separate classes.
Socioeconomic status in interaction with age	12 classes based on total household income interacted with age brackets.
Source of income/education in interaction with age	36 classes derived from source of income or education in interaction with age.
Zip code clusters for somatic care	10 clusters of four-digit zip codes (which represent a village or town or parts of either) based on regional characteristics.
Household size in interaction with age	13 classes based on the number of residents per street address in interaction with age.
Pharmacy-based cost groups (PCG)	43 classes based on individual usage of pharmaceuticals prescribed in year *t-1*. Individuals qualify for a PCG when they exceed a predefined threshold of daily doses of linked pharmaceuticals.
Diagnosis-based cost groups (DCG)	27 classes based on clusters of specific hospital inpatient and outpatient diagnoses in year *t-1*.
Durable medical equipment cost groups (DMECG)	15 classes based on the use of durable medical equipment in year *t-1* related to specific chronic conditions.
Physiotherapy diagnosis groups (PDG)	5 classes based on specific diagnoses from physiotherapy visits in year *t-1*.
Multiple-year high-cost groups	9 classes based on high health care spending on somatic care in years *t-1*, *t-2* and *t-3*. For instance, spending in the top 0.5% in all three years is one of the classes.
Prior spending on home care	10 classes based on cumulative spending on homecare in years *t-1*, *t-2* and *t-3* combined.
Historical somatic morbidity	2 classes based on whether the insured individual in year *t*-3 qualified for at least one of the six morbidity adjusters listed above.
Multiple-year pharmaceutical costs	2 classes based on whether the insured individual had pharmaceutical spending in the top 25% in year *t-1*, *t-2* and/or *t-3*.

6.3.2 Data

Table 6.1 describes the risk adjusters that make up the Dutch RE model for somatic care in 2022.[6] In addition to the dataset that is used to calibrate this model, we use the RE data for the models for two additional, preceding years: the RE models for 2020 and 2021, based on the enrolled population in 2017 and 2018, respectively. In terms of risk adjusters and risk classes, the models for 2020 and 2021 differ only modestly[7] from the 2022 model. Through these datasets, we have access to the individual-level health spending from 2017 to 2019 and the risk adjusters used in the RE models of 2020 to 2022. By combining the data of these three years, we simulate a prediction model for health spending per individual that could be used by insurers to derive risk-rated premiums.[8] In reality, insurers will have different individual-level information available than used in this analysis. For instance, competing insurers will have more in-depth information on the actual health utilization of insured than that used in this study but they will only have that information about those already insured with them. Still, for this simulation, the extensive, available information facilitates an interpretation of theoretical premiums. In addition to the premiums, we replicate the Dutch RE model for 2022 to derive risk-adjusted subsidies. By applying the equations of the previous section, we identify the realized proportion of cross-subsidies. Below, we describe our methodology in more detail.

6.3.3 Methods

In the simulation analysis, we distinguish between a 'prediction model', a 'subsidy model', and a 'premium model'. With these three models we can calculate our primary measure of interest: the cross-subsidies realized by RE as a proportion of the complete cross-subsidies. The *prediction* model is used to estimate individual-level health spending based on information available to insurers. The *subsidy* model is based on the Dutch RE model for 2022 to find individual-level risk-adjusted subsidies to insurers. The *premium* model serves to combine the outcomes of the first two models to find risk-rated premiums in scenarios with and without subsidies. Table 6.2 provides an overview of these models before the detailed description in the next sections.

6 The risk equalization system includes separate models for somatic care, mental care, and out-of-pocket spending under the mandatory deductible. In this paper we focus on the former, which covers roughly 90 percent of total spending under the basic health insurance.

7 The Dutch RE model for somatic care for 2020 had two separate classifications for DCGs, one with sixteen primary DCGs and one with eight secondary DCGs. The models for 2021 and 2022 had one combined classification with 27 DCGs. Also, the model for 2022 has 43 PCGs, while the models for 2020 and 2021 had 38 and 39, respectively. Moreover, the 2022 model has fifteen DMECGs, which were eleven in the 2020 model and increased to fifteen for the 2021 model. Last, the Multiple-year pharmaceutical costs and Historical somatic morbidity were introduced for the 2022 model.

8 This information is used for the prediction model, and we explicitly assume for this research that this captures all information that is available to insurers to predict health spending.

Table 6.2. Overview of the three models used in the analysis to determine premium levels

Name of the model	Description
Prediction model	A model theoretically used by insurers to calculate a prediction of the individual-level health spending of consumers in the contract year, using multi-year risk adjuster information and prior health spending.
Subsidy model	A replica of the risk equalization model of the Netherlands for 2022 to derive individual-level risk-adjusted subsidies to compensate insurers for the variation in expected health spending of consumers.
Premium model	A way to determine the individual-level premium for consumers: in the absence of risk-adjusted subsidies and premium-rate restrictions, the premium is equal to the result from the prediction model; in the presence of risk-adjusted subsidies, the premium is equal to the result of the prediction model, subtracted by the result of the subsidy model.

6.3.3.1 The prediction model

Using the data described in section 3.2, we simulate an extensive prediction model to predict health spending per individual ($\hat{y}_i$) in 2019. In terms of predictors, the model includes variables based on the individual-level health spending in 2017 and 2018, as well as all the risk adjuster information available in the described datasets from Dutch RE.[9] For instance, the Pharmacy-based cost groups, based on pharmaceutical usage in year *t-1* for the respective RE model, are available for three years. Thus, to predict health spending in 2019, pharmaceutical usage from 2016 through 2018 is included.

By combining the three years of data, the total number of predictors stretches beyond 600. Many of these predictors will actually overlap. For instance, the predictors based on the age and sex of an individual can only change once over the three years included (five-year age brackets are used). To filter the selection of predictors to include those that make a statistically significant contribution to the predictive performance of the prediction model, we apply a stepwise regression method. This simple and interpretable model iteratively adds explanatory variables to the regression model and removes those where the statistical significance is below the defined threshold.[10] We use this method to compute a prediction model that effectively applies the available information to predict

9 Without assumption 4 in Section 2 of this paper, insurers will not have access to all this information. Moreover, individuals that are new enrollees for an insurer will bring little historical data for the insurer to use in their risk classification. This may not be a problem if the switching rate is low and almost all switchers have low expected costs.

10 For the stepwise regression, we use an entry level of 0.005 and a stay level of 0.01. The entry level specifies that a variable must be significant at the 0.005 level before it can be entered into the model, while the stay level determines that a variable in the model must be significant at the 0.01 level for it to remain in the model.

health spending in 2019. Given the typical lognormal distribution of health spending, the resulting estimations are expected to resemble Figure 6.1, depending on the quality of the prediction model. The estimations of expected health spending $\hat{y}_i$ are used to derive premiums, as will be discussed in section 6.3.3.3.

6.3.3.2 The subsidy model

To calculate the risk-adjusted subsidies (s_i) on an individual level, we replicate the Dutch RE model of 2022. Using the circa 200 risk classes included in the RE model of 2022 (Table 6.1), we predict health spending in 2019 through an OLS regression. For this model, we exclusively use the information from the RE model for 2022, so the RE information from the additional years used for the prediction model (discussed in section 6.3.3.1) is not included. Since year *t* refers to 2019, the information that underlies the risk adjusters included in the model is derived from 2019 and before (see Table 6.1 for an overview of the years the information stems from).

In our analysis, the resulting predictions of individual-level health spending by the RE model of 2022, REP_i, feed into an internal subsidy model. More specifically, the subsidy s for individual i is calculated according to Equation 6.3.

6.3.3.3 The premium model

In the absence of subsidies, the risk-rated premium p_i equals $\hat{y}_i$. We use Equation 6.2 to derive cross-subsidies (SCS) in a situation of complete cross-subsidization.

To find the risk-rated premiums in the scenario with risk-adjusted subsidies ($p_{i,sub}$), we use Equation 6.4 to subtract the derived risk-adjusted subsidies s_i from the predictions of individual-level health spending $\hat{y}_i$ (used as risk-rated premiums in the absence of subsidies).

6.3.3.4 Evaluating the impact of risk-adjusted subsidies

To calculate the extent to which the Dutch RE model of 2022 achieves the cross-subsidies that would be achieved by community-rating across the market (RPCS), we apply Equation 6.5. This provides a direct answer to the central research question. To provide a deeper interpretation of the effect of risk-adjusted subsidies, we supplement the RPCS with some additional outcome measures. We explore the variation in premiums for both scenarios through quantifying the share of premiums that exceed specific thresholds or fall outside a specific bandwidth. These measures indicate the inequality in premiums that remains after applying risk-adjusted subsidies.

6.4 RESULTS

This section presents the results of our empirical analysis. First, a range of descriptive statistics are shown for the datasets used before we present the results of our simulation analysis.

6.4.1 Descriptive statistics

The three sets of data used for this research differ slightly in terms of size and statistics. Table 6.3 provides descriptive statistics of the (combined) datasets. For the three consecutive years, the data show an upward trend both in terms of spending and in terms of the number of insured individuals. Throughout the analysis, we followed the procedure

Table 6.3. Descriptive statistics of the datasets used in this study

	2017	2018	2019	Present in all three years
Number of individuals	17,128,725	17,256,295	17,365,571	16,651,263
Insured years	16,839,948	16,951,700	17,041,789	16,514,947
Mean spending	€ 2,355	€ 2,408	€ 2,487	€ 2,492
Standard deviation of spending	€8.452	€8.622	€8,813	€8.569
Frequencies				
Women	50.6%	50.5%	50.5%	50.6%
Ages 0 to 17	20.0%	19.7%	19.5%	18.3%
Ages 18 to 34	20.6%	20.8%	21.0%	20.8%
Ages 35 to 44	12.1%	11.9%	11.9%	11.9%
Ages 45 to 54	15.0%	14.8%	14.4%	14.7%
Ages 55 to 64	13.4%	13.5%	13.7%	14.0%
Ages 65 to 74	11.0%	11.1%	11.3%	11.6%
Ages 75 and older	8.0%	8.1%	8.4%	8.6%
PCGs > 0	16.7%	16.8%	17.2%	17.7%
DCGs > 0	10.5%	11.4%	11.7%	12.0%
DMECGs > 0	3.8%	4.5%	4.6%	4.7%
PDGs > 0	1.9%	2.0%	2.8%	2.8%
Multiple-year high-cost groups > 0	6.0%	6.1%	6.2%	6.4%
Home care spending groups > 0	2.4%	2.4%	2.4%	2.5%
At least in one of the above morbidity classes	24.6%	25.2%	25.6%	26.2%

Note: the abbreviations used refer to Pharmacy-based cost groups (PCGs), Diagnosis-based cost groups (DCGs), Durable Medical Equipment cost groups (DMECGs) and Physiotherapy diagnosis groups (PDGs).

of annualizing spending and weighing the outcomes with the fraction of the year the individual was enrolled in health insurance.[11] To prevent concerns over missing values, we only included individuals present in all three years.[12] 16,651,263 individuals remained within the combined dataset. The mean health spending in 2019 for the individuals present in all three years is €2,492, slightly above the average for the entire population in 2019.

6.4.2 Prediction of health care spending $\widehat{Y_I}$

For all 16.7 million individuals in the final dataset, we used the prediction model to estimate their expected health spending in 2019. In addition to the roughly 600 RE risk classes that were available for the three years, we used the level of health spending in 2017 and 2018 as predictors. For instance, individuals with health spending between €17,500 and €30,000 in 2018 were allocated to a risk class, while those with spending above €250,000 were allocated to another. In Table A6.1 in the Appendix, the allocation of individuals to risk classes based on historical health spending is shown.

After applying the stepwise regression method for selecting variables of statistical significance, 401 variables remained in the prediction model. Table A6.2 in the Appendix shows the twenty risk classes in the prediction model with the largest (absolute) coefficient. The list predominantly consists of the risk classes based on the highest amount of health spending in the preceding year (2018) and the risk classes designed to detect individuals that used (extremely) costly pharmaceuticals.

6.4.3 Risk-rated premiums in the absence of risk-adjusted subsidies

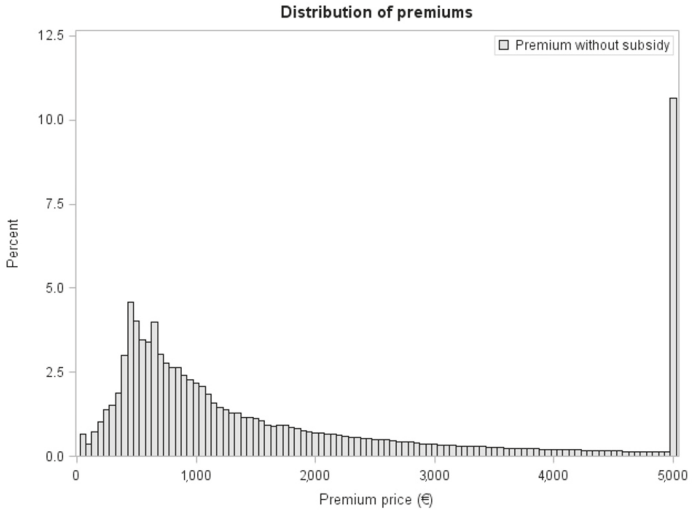
In this scenario the individual-level predictions from the prediction model are used directly as the risk-rated premiums. Figure 6.3 illustrates the distribution of premiums that result from our prediction model with 401 explanatory variables, in the absence of any subsidies. For ease of presentation, the extremes of the figure have been set to €0 and €5,000.[13] Therefore, all individual-level premiums that exceed either threshold are, only for the purpose of visualization within the figure, set to that respective level. The average premium is equal to the mean level of health spending (€2,492). However, most individuals (76.6%) in this scenario of complete risk-rating are charged a premium below that average. Still, ten percent of the population is charged a premium above €5,000, i.e., higher than twice the average premium, and the highest premium is €656,173. Furthermore, the

11 Individuals may, for specific reasons, have stopped their health insurance at any point during the year (e.g., due to emigration) or have enrolled after January 1st. Therefore, individual-level spending is annualized and weighted by the fraction of the year an individual was enrolled. For example, spending accumulated by individuals who were enrolled for half of the year is doubled and attributed a weight of 0.5 in the analyses.

12 For individuals present in 2022 but absent in any of the prior years, information is missing that could help predict their health spending in 2022. In reality these individuals would nevertheless require an estimation of their health spending for insurers to define a premium level.

13 The linear regression model generates some negative predictions for health spending in 2019, which is a result of the skewed distribution of health spending in the population and multicollinearity in the applied risk adjusters within the model.

Figure 6.3. Distribution of premiums in the scenario of risk-rating in the absence of risk-adjusted subsidies



Note: Within this figure, premiums below €0 or above €5,000 are set to those respective values for the purpose of visualization of the results.

variation is quantified through the standard deviation: €5,479. Compared to the average premium, this standard deviation indicates a substantial variation in premiums. By applying Equation 6.2, we derived the mean absolute deviation from the average premium to

Table 6.4. The distribution of premiums in a scenario of risk-rated premiums without risk-adjusted subsidies

		Without subsidies
Premium above average (€2,492)		23.4%
Premium above €5,000		10.3%
Premium above €10,000		4.2%
Premium above €20,000		1.4%
Maximum premium		€ 656,173
Premium within a bandwidth of +/- €250 around average		5.2%
	Below bandwidth	73.9%
	Above bandwidth	21.0%
Premium within a bandwidth of +/- €500 around average		10.4%
	Below bandwidth	70.6%
	Above bandwidth	19.0%
Premium within a bandwidth of +/- €1,000 around average		22.4%
	Below bandwidth	61.7%
	Above bandwidth	15.9%

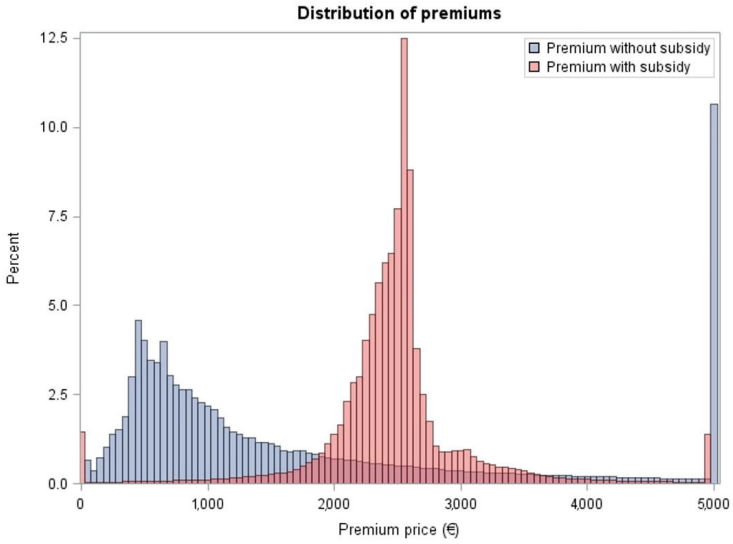
quantify the average level of cross-subsidization per individual that is required to achieve complete cross-subsidization: €2,338.

Table 6.4 presents more information on the distribution of premiums in this scenario. Although over three-quarters of the population are charged a below-average premium, an important share of the population is charged a substantial premium that is over double the average premium. Moreover, relatively many individuals are charged a premium outside the selected bandwidths around the average premium (€500, €1,000, and €2,000).

6.4.4 Risk-rated premiums in the presence of risk-adjusted subsidies

In addition to the prediction model, we simulated a subsidy model to simulate the second scenario: risk-rating in the presence of risk-adjusted subsidies. After replicating the Dutch RE model for 2022 to find the individual-level subsidies, we subtracted the subsidies from the premiums derived from the prediction model in the previous scenario. In Figure 6.4, the distribution of premiums after applying the risk-adjusted subsidies is shown. Compared with the previous scenario (risk-rated premiums in the absence of subsidies), the distribution has shifted towards the average as the subsidies compress the premium levels. Relatedly, the standard deviation of the premiums reduces from €5,479 in the previous scenario to €1,568 in this scenario.

Figure 6.4. Distribution of premiums in the scenario of risk-rating in the presence of risk-adjusted subsidies (red), compared to a scenario of risk-rating without subsidies (blue).



Note: Within this figure, premiums below €0 or above €5,000 are set to those respective values for the purpose of visualization of the results.

Crucially for this research, the mean absolute deviation of premiums from the average is reduced to €454. By applying the findings from the current and previous scenarios to Equation 6.5, we found a value of 0.806 for the RPCS. Through that number, the main research question can be answered subject to the assumptions made in Section 6.2: the cross-subsidies that are created through the risk-adjusted subsidies equal 81% of the cross-subsidies that would be achieved by community-rating across the market.

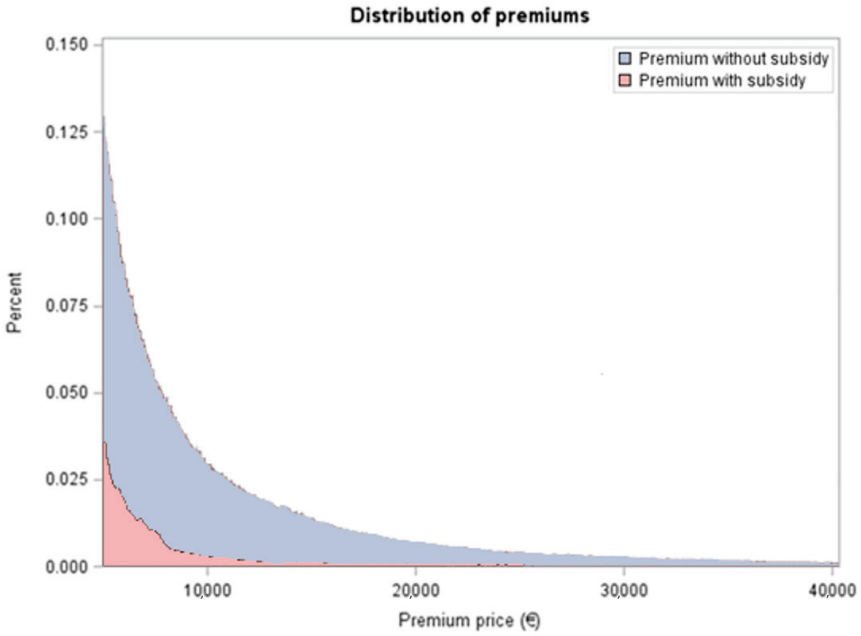
In Table 6.5, we show further indications of the impact of risk-adjusted subsidies. Although a larger proportion of individuals is charged a premium above the average premium, the share of individuals with substantial premiums (above €5,000) is significantly reduced. Moreover, after the introduction of risk-adjusted subsidies, the share of individuals charged a premium within the defined bandwidths around the average strongly increases.

Table 6.5. The distribution of premiums in a scenario of risk-rated premiums 1) without risk-adjusted subsidies and 2) with risk-adjusted subsidies.

	Without subsidies	With subsidies
Premium above average (€2,492)	23.4%	49.9%
Premium above €5,000	10.3%	1.3%
Premium above €10,000	4.2%	0.25%
Premium above €20,000	1.4%	0.071%
Maximum premium	€ 656,173	€ 157,132
Premium within a bandwidth of +/- €250 around average	5.2%	61.9%
Below bandwidth	73.9%	22.6%
Above bandwidth	21.0%	15.5%
Premium within a bandwidth of +/- €500 around average	10.4%	79.7%
Below bandwidth	70.6%	10.2%
Above bandwidth	19.0%	10.5%
Premium within a bandwidth of +/- €1,000 around average	22.4%	90.8%
Below bandwidth	61.7%	4.5%
Above bandwidth	15.9%	4.7%
% of cross-subsidies realized	*n/a*	80.6

To grasp the variation of premiums above average, Figure 6.5 illustrates the distribution of premiums that are exclusively over €5,000 for both scenarios (i.e., the stacked bars on the right-sided extremes in Figures 6.3 and 6.4 As shown before in Table 6.5, the share of individuals charged a premium above that threshold is substantially reduced by the risk-adjusted subsidies: from 10.3% to 1.3%. Moreover, the figure demonstrates the spread of these premiums, illustrating how the subsidies converge the premiums towards the average (of €2,492).

Figure 6.5. Distribution of premiums above €5,000 in the scenario of risk-rating with risk-adjusted subsidies (red) and without subsidies (blue)



6.4.5 Risk-rated premiums in the presence of risk-adjusted subsidies and outlier risk-sharing

In practice, regulators of social health insurance markets often implement some form(s) of ex-post risk-sharing as a supplement to RE. To get an impression of the potential effect of risk-sharing on the leftover variation in risk-rated premiums, we simulated a modality of outlier risk-sharing (ORS). Here, insurers are compensated for (a portion of) health spending by consumers above a defined threshold (McGuire & Van Kleef, 2018b). To examine how ORS impacts premiums in a risk-rated system, we simulated two models of ORS as additional subsidies to the risk-adjusted subsidies. The two variants have different thresholds to cut off spending: the first ORS variant cuts off 100% of spending above €100,000 (which applies to 0.20% of the population), and the second does so at €25,000 (applies to 1.9% of the population). For the application of ORS, we first removed individual health spending above the defined threshold and then re-estimated the premium model on the leftover spending. After, we re-estimated the model for the risk-adjusted subsidies on the leftover spending and defined the new premiums.

Table 6.6 presents the effects of the ORS modalities on the distribution of premiums. Since the ORS threshold applies for only a small proportion of the population, the number of

Table 6.6. The distribution of premiums in four scenarios

	Without subsidies	With subsidies	With subsidies and ORS €100,000	With subsidies and ORS €25,000
Premium above average (€2,492)	23.4%	49.9%	49.4%	49.9%
Premium above €5,000	10.3%	1.3%	1.3%	0.79%
Premium above €10,000	4.2%	0.25%	0.24%	0.00014%
Premium above €20,000	1.4%	0.071%	0.050%	*n/a*
Maximum premium	€ 656,173	€ 157,132	€ 44,980	€ 11,343
Premium within a bandwidth of +/- €250 around average	5.2%	61.9%	62.3%	62.4%
Below bandwidth	73.9%	22.6%	22.6%	23.1%
Above bandwidth	21.0%	15.5%	15.1%	14.5%
Premium within a bandwidth of +/- €500 around average	10.4%	79.7%	79.7%	81.3%
Below bandwidth	70.6%	10.2%	10.1%	9.3%
Above bandwidth	19.0%	10.5%	10.2%	9.4%
Premium within a bandwidth of +/- €1,000 around average	22.4%	90.8%	91.3%	93.9%
Below bandwidth	61.7%	4.5%	4.2%	2.2%
Above bandwidth	15.9%	4.7%	4.5%	4.0%
% of cross-subsidies realized	*n/a*	80.6	82.0	86.1

Note: The four scenarios referred to in the title of the table are a scenario of risk-rated premiums, 1) without risk-adjusted subsidies, 2) with risk-adjusted subsidies, 3) with risk-adjusted subsidies and outlier risk-sharing for spending above €100,000, and 4) with risk-adjusted subsidies and outlier risk-sharing for spending above €25,000.

individuals that directly benefit is relatively limited. Nevertheless, the effect is clear: premiums converge further. Moreover, applying ORS increases the realized cross-subsidies as a proportion of the cross-subsidies that are realized through community-rating across the market: the ORS scenarios increase the RPCS from 80.6% to 82.0% and 86.1%, respectively. The risk-adjusted subsidies create an important portion of cross-subsidies that are realized through community-rating and the addition of ORS leads to stronger cross-subsidization. Still, after applying a sizeable ORS modality with a threshold of €25,000, 13.9% of the 'complete cross-subsidization' (Equation 6.2) is not realized. In the next and final section, we discuss the implications of our findings.

6.5 DISCUSSION AND CONCLUSION

In this paper we explored the extent to which a state-of-the-art RE model creates cross-subsidies on a competitive health insurance market. Through an extensive prediction model, based on information on health spending and risk adjusters from three prior years, we calculated individual-level risk-rated premiums in a scenario where subsidies and premium-rate restrictions are absent. In a second scenario, we supplemented these premiums with risk-adjusted subsidies through a simulation of the Dutch RE model of 2022. The substantial variation present in the first scenario of risk-rated premiums is significantly reduced through the addition of subsidies in the second scenario, which leads premiums to converge towards the average. The share of premiums that exceed defined thresholds or fall outside selected bandwidths around the average is significantly reduced. Overall, the risk-adjusted subsidies create 81% of the cross-subsidies that would be achieved by implementing community-rating across the market, subject to the restrictive assumptions made in this paper. Additionally, we simulated the effects of outlier risk-sharing as a supplement to the risk-adjusted subsidies. Outlier risk-sharing with a threshold of 25,000 euros further reduces the variation in premiums and increases the percentage of achieved cross-subsidies from 81% to 86%.

6.5.1 Assumptions

The findings of this study are subject to important assumptions, as outlined in the second section of this paper. First, we assumed insurance contracts to cover the annual health care spending of individuals, rather than that of groups or families. Without this assumption, the price for risk-rated premiums for multi-person contracts are different but could theoretically become a simple summation of the combined individual premiums. Second, we assumed the presence of a standardized benefit package, a strong insurance mandate, and open enrollment. The absence of these assumed regulatory interventions has significant implications for the findings of this study. Without a standardized benefit package, differentiation in the insurance packages offered by insurers may result in adverse selection[14], which leads to price variation in addition to the expected health care spending. Moreover, without an insurance mandate and open enrollment, the insurance market could eventually collapse. Individuals who reflect significant risk to insurers can be rejected from insurance coverage, while healthy individuals may choose not to insure themselves. As a result, the risk-adjusted subsidies used in our simulation may not be sufficiently funded through limited insurance uptake and contribution to the subsidy model. A third assumption made in our study is that insurers charge risk-rated premiums based on the *expected health care costs* per individual and do not include any type of *loading fees*. In practice, premiums for insurance include a loading fee that reflects administra-

14 Adverse selection may occur as the result of information asymmetry between the consumer and the insurer: the consumer has more information about her health status and expected health care expenses than the information that the insurer has used for risk-rating the premiums.

tive costs, transaction costs for determining individual-level risk, capital costs, and other factors that alter the effective premium. If regulators do not include the loading fee in the risk equalization model, which is not done in the Netherlands, this will lead to different outcomes than those presented in our research. Moreover, for practical reasons, e.g., administrative costs and transparency of the premiums, insurers may not include all risk factors, which may also lead to different outcomes than those presented in our research.

Fourth, we assume that insurers have information about *all* ex-ante risk characteristics of *all* individuals who apply for insurance and use that to derive risk-rated premiums. This is certainly different from reality. Competing health insurers will have information at a more detailed level on health utilization and spending based on declarations than the risk adjuster information we used. However, insurers will only have this specific information available for their current enrollees, which limits the data for the model and obscures the calculation of an adequate premium for new enrollees. This could significantly alter the outcomes of our study in terms of resulting premiums. Consumers could strategically switch insurers on an annual basis to acquire coverage for a below-market premium that fails to identify the true risk of the individual. This risk for insurers could result in a loading on top of the premium for new enrollees, impacting the results of our study.

Moreover, the 81% we found in this study as the realized proportion of cross-subsidies achieved by RE is subjective to the quality of the model used for premium-setting and the model to calculate risk-adjusted subsidies. Therefore, it remains an empirical question whether, in reality, insurers are able to apply a more advanced model for premium-differentiation, which would lower the found proportion. On a free market, insurers could have applicants fill out extensive surveys on health status to derive premium levels through the new information. Alternatively, the regulator could implement a mandate for the insurers to share the acquired information on a shared platform to avoid the phenomenon of adverse selection by switching consumers. If all information is available to all insurers, the results in that scenario should not be too different from our study.

As a fifth assumption, we described that both the competing insurers and the regulator use OLS regression to estimate the individual-level health spending. However, in reality insurers may exploit the gathered information through more precise and complex methods than our relatively simple regression. Also, the methods used between the various insurers will likely deviate. As a result, their models will likely lead to different predictions of health spending (and risk-rated premiums) and a different level of cross-subsidies that would be achieved through community-rating across the market. So, the degree of cross-subsidies that is realized through risk-adjusted subsidies depends on both the subsidy model and the prediction model. Further research could use different information and alternative methods to predict health spending.

For the subsidy model, we used the Dutch RE model for somatic care for 2022. The Dutch system, however, also includes a model for mental care and one for out-of-pocket spending under the mandatory deductible. Moreover, the model has been expanded since 2022, both in terms of risk adjusters and in terms of the estimation method applied.[15] Therefore, while the findings are relevant for the Netherlands, the use of more recent models may lead to slightly different results. Extrapolating the findings to international health insurance systems with comparatively advanced RE models requires further research. Still, like our findings, such models are expected to achieve a large proportion of the cross-subsidies that would be achieved through community-rating across the market.

6.5.2 Affordability and equality of premiums

Despite the implications of the assumptions made, our study demonstrates that a state-of-the-art RE model achieves a substantial share of the cross-subsidies that are realized through community-rating across the market. After applying risk-adjusted subsidies 1.3% of premiums are above €5,000, and 90.8% fall within the bandwidth of €1,000 below and above the average premium. However, from the perspective of *affordability and equality of premiums*, the leftover variation may still be problematic for regulators and for (very) high-risk individuals. Also, there is the reclassification-risk for individuals to become a (very) high financial risk to insurers in the future, threatening their affordability of future coverage. The advantage of this reclassification-risk, however, is that individuals are expected to become more (financially) aware of the benefits of prevention and healthy behavior. This could be an important effect with the growing need for health services in populations.

If affordability of coverage is indeed considered problematic, policymakers could consider additional instruments to further reduce variation in premiums. Outlier risk-sharing – as applied in our simulation analysis – is one example of such an additional instrument. An alternative instrument could be high-risk pooling (HRP), which allocates high-risk individuals to a compensation-pool for their expected high level of spending. In recent studies, Withagen-Kosten et al. (2024) and Oskam et al. (2025) used a random forest algorithm to identify groups of insured individuals of 'high-risk'. They then used HRP to specifically compensate these individuals. Using the results of our premium model, individuals that are charged the highest premiums (i.e., the individuals in the furthest right-sided bars in Figure 6.4) could be considered as high-risk individuals for whom the insurer receives a cost-based compensation through HRP. Alternatively, if premium-rate restrictions are to be partially relaxed to ensure all premiums are constrained within a defined bandwidth in order to maintain affordability, risk-sharing could also be applied for premiums that fall outside the agreed bandwidth. Within the bandwidth, a limited number of premium levels could be allowed to protect transparency on the market. Irrespective of the modality chosen, risk-sharing has an important downside: incentives for efficiency for insurers are

15 Since 2024, the Dutch RE model for somatic care is estimated through a constrained regression as an expansion to the OLS regression.

reduced through the additional cost-based compensation they receive (Van Kleef et al., 2018b). Therefore, improving affordability and equality of premiums through risk-sharing will reduce the incentives for efficiency. Hence, there is a tradeoff between resulting incentives for risk selection stemming from combining community-rating with RE and reduced incentives for efficiency (and transparency) stemming from combining RE with risk-sharing.

Another option to create stronger cross-subsidization between the risk types on the market would be to further improve the subsidy model. However, provided the RE model studied in this research is already considered state-of-the-art quality, the question remains to what extent the model can be further improved. The goal is then to minimize the misfit between the estimations by the subsidy model and the prediction model used by insurers for the premium-setting. Insurers could be mandated to share their prediction models with the regulator to improve the future subsidy model. To gain a financial edge over competitors and the subsidy model, insurers are then incentivized to find additional predictors, which will in due time subsequently improve the subsidy model. However, if insurers are mandated to immediately report these additional predictors, this may effectively discourage investment into finding them. Moreover, it remains to be seen whether a prediction model can ever be sophisticated enough to adequately predict the health spending of (high-risk) individuals.

In the case that proposed strategies to further improve cross-subsidies are deemed unsuitable, regulators could resort to premium-rate restrictions. Complete community-rating ensures equality of premiums but also creates regulation-induced selection incentives that may threaten the functioning of the health system. The size of these incentives depends, for an important part, on the quality of the present RE system. Moreover, in a valuation of strategies, the threat of the regulation-induced selection incentives in such a scenario should be weighed against the loss of efficiency incentives if risk-sharing is chosen to reduce premium variation in the absence of rate-restrictions. An advantage of community-rated premiums is that insurance products, and their respective premiums, are more easily compared by consumers, improving competition in the market and its efficiency as a result.

A moderation of premium regulations could be an interesting alternative: a regulated bandwidth for premiums could (largely) protect equality of premiums, while an important share of regulation-induced selection incentives is mitigated. As shown in our simulation, most premiums fall between moderate bandwidths after applying risk-adjusted subsidies. Applying a premium bandwidth does decrease transparency in the market of offered insurance products and premiums, impeding simple comparisons between products for consumers. As a result, competition could be hindered, limiting the desired increase to efficiency on the market. Therefore, as stated before, the key tradeoff in this conundrum remains between the threat of regulation-induced selection incentives and efficiency.

6.5.3 Conclusion

The starting point of our simulation was that regulation-induced selection incentives are a problem for the functioning of a competitive health insurance market. Our study demonstrates that risk-adjusted subsidies realize a notable share of the cross-subsidies that are achieved through community-rating. Specifically, the state-of-the-art RE model that we analyzed achieves 81% of the cross-subsidization that would be attained by community-rating across the market, subject to our assumptions. Forms of ex-post risk-sharing further increase this percentage. However, despite this significant achievement, a variation in premiums remains, necessitating further consideration. The extent to which the remaining variation is acceptable and worthwhile depends on the goals of the regulator and society regarding affordability, equality, and accessibility of premiums, as well as their stance on risk selection by insurers and the functioning of the competitive insurance market. If achieving absolute equality of premiums is of great importance, premium regulation is still necessary. The benefit is that comparison between insurance products and premiums is simplified, improving the competitive aspect of the market and efficiency as a result. However, if regulators are willing to tolerate some variation in premiums to avoid the negative impacts of regulation-induced selection, the current RE models create notable cross-subsidization as subsidy models. While the achieved removal of selection incentives is of crucial importance, further improvement of the model or additional policy measures would be required to address the potential threat to affordability if premium-rate restrictions were to be completely removed.

6.6 APPENDIX

Table A6.1. Overview of the created risk classes based on health spending in 2017 and 2018 that are used in the premium model

Subgroup	Health spending in 2017 (€)	Health spending in 2018 (€)
1	100 and below	100 and below
2	101 to 1,900	101 to 1,900
3	1,901 to 4,500	1,901 to 4,500
4	4,501 to 10,000	4,501 to 8,000
5	10,001 to 35,000	8,001 to 9,000
6	35,001 to 50,000	9,001 to 14,000
7	50,001 to 60,000	14,001 to 17,500
8	60,001 to 100,000	17,501 to 30,000
9	100,001 to 125,000	30,000 to 35,000
10	125,001 to 150,000	35,000 to 50,000
11	150,001 to 250,000	50,001 to 60,000
12	Above 250,000	60,001 to 75,000
13	*n/a*	75,001 to 100,000
14	*n/a*	100,001 to 125,000
15	*n/a*	125,001 to 150,000
16	*n/a*	150,001 to 250,000
17	*n/a*	Above 250,000

Note: The subgroups are created based on the residual spending per cluster of prior health spending. The residual spending results from replicating the RE model of 2022 and subtracting that from the actual health spending of individuals. In doing so, homogeneity within the resulting subgroups is sought with respect to residual spending.

Table A6.2. The twenty risk classes used in the premium with the largest (absolute) coefficient

Risk class	Coefficient
Pharmacy-based cost group #42 in 2019	€ 412,166
Pharmacy-based cost group #41 in 2019	€ 205,414
Health spending in 2018 above € 250,000	€ 137,026
Pharmacy-based cost group #40 in 2019	€ 76,551
Pharmacy-based cost group #37 in 2017	€ 74,202
Pharmacy-based cost group #38 in 2018	€ -71,705
Health spending in 2018 between € 150,001 and € 250,000	€ 62,830
Pharmacy-based cost group #37 in 2018	€ -54,857
Health spending in 2018 between € 125,001 and € 150,000	€ 47,415
Diagnosis-based cost group #25 in 2019	€ 44,399
Prior spending on home care #9 in 2019	€ 40,142
Health spending in 2018 between € 100,001 and € 125,000	€ 35,482
Diagnosis-based cost group #24 in 2019	€ 33,169
Pharmacy-based cost group #36 in 2018	€ 32,276
Health spending in 2018 between € 75,001 and € 100,000	€ 25,708
Diagnosis-based cost group #23 in 2018	€ -24,061
Prior spending on home care #8 in 2019	€ 23,500
Diagnosis-based cost group #26 in 2019	€ 23,436
Multiple-year high-cost groups #8 in 2019	€ 23,304
Diagnosis-based cost group #24 in 2018	€ -21,818

7

CONCLUSION AND DISCUSSION

Within this dissertation, new evaluation measures for selection incentives that remain after risk equalization in social health insurance markets were developed, and strategies to further reduce those incentives have been tested. In chapters two through six, the respective contributions of the studies to that goal are described by answering five separate research questions. The first two of these questions relate to the use of innovative methods to evaluate selection incentives in health insurance systems with risk equalization. Research questions three, four, and five concern the simulation of three separate strategies to further reduce selection incentives. The implications of the findings presented within this dissertation are further discussed after the general conclusions of the separate chapters.

7.1 CONCLUSIONS

Sections 7.1.1 and 7.1.2 discuss the findings presented in each of the five chapters through answering the research question posed. Therefore, these sections serve to create an overarching view of the individual contributions of the respective studies to the body of academic literature on risk equalization in health insurance.

7.1.1 Evaluating risk equalization through innovative methods

Question 1: Does health-based prospective risk equalization adequately compensate insurers for individuals diagnosed with a new chronic disease?

In chapter two, the performance of the Dutch (prospective) risk equalization model was investigated through an evaluation of the insurer compensation for individuals newly diagnosed with a chronic disease. The main objective was to assess insurers' incentives for risk selection by estimating the financial outcomes after risk equalization for insurers and the insurer-switching behavior of individuals within the years surrounding diagnosis. Through the multi-year diagnostic information acquired from electronic patient records from general practitioners, individuals diagnosed with a disease for the first time were identified for a variety of chronic diseases. To do so, the presence of (a specific) chronic disease(s) was compared for subsequent years to find the year of diagnosis. For the identified subgroups of individuals diagnosed with a new chronic disease in year t, the evolution in health spending, risk equalization payments, and rates of insurer-switching were tracked for the period from *t-2* to *t+2*. The findings reveal a significant payment shortfall in the year of diagnosis. The increase in health spending with the incidence of the disease is not fully compensated by the risk equalization model due to its prospective nature. In the year surrounding diagnosis, the identified subgroups are typically overcompensated in the years prior to diagnosis and undercompensated in years after. However, between the subgroups, the results vary: individuals diagnosed with a chronic disease for the first time are more accurately compensated in the second year after diagnosis than their

relatively sicker counterparts that were already diagnosed with a different chronic disease before. Moreover, a higher propensity to switch insurer was found for individuals diagnosed with a chronic disease for the first time, compared to the general population, but no substantially disproportionate levels of switching were found for specific chronic diseases studied. Subject to various limitations, the general implication of this research is that consumers do, to some extent, respond to their change in health status, which means that the observed payment gaps could lead to selection incentives for insurers. This suggests that the risk equalization model needs further improvement.

Question 2: What does the heteroscedasticity of residual spending look like, and how can it affect selection incentives?

Chapter three examined the heteroscedasticity of residual spending after risk equalization in health insurance markets, questioning its potential impact on selection incentives. The core objective was to explore how the variation in the financial result within and between these groups can lead to incentives for risk selection, even when risk equalization leads to an average financial result of zero for groups of insured individuals. After replicating the risk equalization model for the Netherlands, the variation in residual spending was quantified to demonstrate the financial uncertainty to insurers across a variety of risk groups. Through simulations influenced by financial investment theory and European capital requirements, the uncertainty in financial outcomes was measured by quantifying theoretical mark-ups on the premiums to accommodate these factors. For risk groups with higher variability in residual spending, the loading on the premiums is likely to be larger. In social health insurance markets where differentiation in premiums is prohibited, the variation in desired mark-ups across subgroups serves as a quantification of incentives for risk selection. Between the risk groups studied in the research, substantial differences in variation in residual spending were found. Typically, variation in residual spending is larger for risk groups with high average health spending than for risk groups with low average health spending, despite both groups having an equal average residual spending of zero. Although the magnitude of the selection incentives that result from heteroscedasticity of residual spending is uncertain, the results of the simulation indicate that these incentives are nontrivial. Therefore, policy measures might be required to effectively reduce selection incentives resulting from differences in variation in residual spending across subgroups.

7.1.2 Simulating strategies to mitigate selection incentives

Question 3: To what extent do modifications to the Dutch DCG-classification reduce selection incentives?

In chapter four, a typical strategy for model improvement was simulated and evaluated: modification of one of the existing risk adjusters of the Dutch risk equalization model.

One of the key risk adjusters to indicate chronic illness, the Diagnosis-based Cost Groups (DCGs), was redesigned to better identify and compensate individuals with a chronic disease. The goal was to better accommodate the heterogeneity of risk of individuals indicated with a DCG and to reduce the persistent predictable losses associated with those individuals. Through utilizing data from Dutch hospital claims covering diagnoses and treatments and applying a new clustering method, the DCG risk adjuster was updated and improved. The redesign included a revision of the underlying hospital diagnoses and treatments that are included in the risk adjuster and the possibility for individuals to acquire more than two DCG flags (in the original DCGs, individuals could only be flagged for a single primary DCG and a single secondary DCG at once). The impact on predictive accuracy of the model and compensation adequacy was compared to the baseline risk equalization model without the DCG alteration. Results indicate substantial reductions in predictable profits and losses, notably improving compensation precision at the level of the hospital treatments that were used to construct the DCGs. Moreover, the fact that individuals can be flagged for multiple DCGs aids in capturing more comprehensive expenditure among individuals with multimorbidity. However, effects are less pronounced for subgroups identified through the external diagnostic details acquired from electronic patient records from general practitioners. Therefore, despite the improvements, individuals diagnosed with a chronic disease remain largely undercompensated by the risk equalization model, reflecting limitations in the recognition and compensation of these individuals by the model. Further improvement of the risk equalization system is therefore of significant importance to enhance fairness, efficiency, and accessibility for these vulnerable groups and the health system entirely.

Question 4: To what extent does high-risk pooling reduce the selection incentives that remain after sophisticated risk equalization?

In chapter five, a different approach for the reduction of selection incentives was simulated through the implementation of a risk-sharing modality called high-risk pooling (HRP). The objective was to study the potential of this approach to mitigate selection incentives within the Dutch health insurance system by specifically targeting high-risk individuals for additional compensation. Unlike traditional risk-sharing modalities, HRP allows insurers to allocate individual consumers to a high-risk pool at the start of an insurance contract to reduce selection incentives. By incorporating historical data on health spending and risk adjuster information in three different predictive models, individuals with a high risk of substantial residual spending were identified and allocated to the high-risk pool. The results of these three models – a prior-year spending model, an ordinary least squares (OLS) regression model, and a more sophisticated Random Forest model – were compared, and the Random Forest model identified the top spenders with greater accuracy. Following this classification, five scenarios for the HRP size were tested (i.e., the top one to top five percent of predicted residual spenders). The findings underscore HRP's

efficacy in reducing selection incentives, presenting a 42% reduction in the undercompensation of individuals with chronic conditions when targeting only the top one percent of residual spenders for HRP, using only 2.6% of the total budget in the risk equalization system. As a result, predictable profits and losses are significantly reduced, alleviating incentives for risk selection for insurers, whilst largely maintaining incentives for cost control. These findings suggest that HRP is a promising tool to supplement imperfect risk equalization. When compared to more commonly applied modalities of risk-sharing, HRP demonstrated far stronger effectiveness at reducing the undercompensation of chronically ill individuals when allocating the same budget for risk-sharing purposes.*Question 5: To what extent does state-of-the-art risk equalization create cross subsidies in a competitive health insurance market?*

In the sixth chapter, another strategy to reduce (or remove) selection incentives was simulated: the relaxation of premium-rate restrictions. Without premium-rate restrictions, regulation-induced selection incentives are absent if premiums reflect the expected spending of individuals. However, in that scenario, premiums can become unaffordable for the individuals with the highest risk of (significant) health spending. In the presence of sophisticated risk equalization, a sizeable proportion of the predictable variance in health spending is compensated. Therefore, in this chapter, a measure was developed to quantify the proportion of cross subsidies that are realized by risk equalization compared to those realized by complete community-rating. For the analysis, a prediction model using historical data on health spending and risk adjuster information was applied to estimate individual-level health spending. Through the predictions, risk-rated premiums were defined to find the level of cross-subsidization that would be achieved by community-rating across the market. By simulating the Dutch risk equalization model, its effectiveness to achieve those cross subsidies was quantified. The results demonstrate that the model achieves 81% of the cross-subsidies that would be achieved by community-rating across the market, given certain assumptions. The payments from the risk equalization model play a pivotal role in diminishing the magnitude of premium differentiation in the risk-rated scenario. Additionally, a simulated modality of risk-sharing, outlier risk-sharing, further diminishes premium disparities and increases the proportion of cross-subsidies achieved. The study demonstrates that the sophisticated Dutch risk equalization model achieves a high level of cross-subsidization but still leaves room for considerable variation in premiums. While the achieved removal of selection incentives is of crucial importance, further improvement of the model or additional policy measures would be required to address the potential threat to affordability if premium-rate restrictions were removed.

7.2 OVERARCHING STRENGTHS AND LIMITATIONS OF THIS RESEARCH

Before discussing the broader implications of the findings presented above, it is worthwhile to reflect on the overarching strengths and limitations of the research throughout this dissertation. Within chapters two to six, the respective strengths and limitations per chapter have already been addressed with a focus on the applied methodology.

In addition, some overarching strengths and limitations are worth mentioning. First, in terms of strengths, the data sources used for the empirical work in this dissertation ensure that the findings present an accurate reflection of reality. The administrative data used to simulate the risk equalization models within the studies are the exact same data as used in practice to estimate the actual Dutch model for the corresponding years that are studied. Therefore, the results provide an accurate reflection of reality in terms of selection incentives after risk equalization.

Furthermore, the diagnostic information from general practitioners acquired from the Nivel Primary Care Database also served a pivotal role within this dissertation to quantify selection incentives for (specific) chronically ill individuals. Both the level of detail and the size of this data source contribute significantly to the relevance of this dissertation. The objective information about the health status of over one million individuals included in the dataset, as diagnosed and registered by general practitioners, enables the identification of specific subgroups within the population as a basis for the evaluation of risk equalization. Through this dataset, an accurate indication is acquired of the presence of specific chronic diseases for all individuals included. This level of detail exceeds that of alternative data sources available.

Therefore, the combination of the two data sources within the chapters of this dissertation leads to valuable insights into the predictable profits and losses insurers face for specific subgroups of the population under the applied risk equalization models. Thereby, this research provides an important contribution to the literature about risk equalization and risk selection incentives.

An important overarching limitation of this dissertation is that the empirical analyses are exclusively based on Dutch data. Therefore, the findings of the dissertation are not directly transferable to health insurance systems in other countries that employ risk equalization. Since the Dutch risk equalization model is regarded as (one of) the best in the world, the findings will likely be different for health insurance markets abroad. Nevertheless, the general implications should still hold and underscore the need for similar research in countries beyond the boundaries of the Netherlands.

Another important limitation is that the findings of this dissertation are subject to timing, as the Dutch risk equalization model is annually updated by adding and/or altering risk adjusters. By doing so, the model is repeatedly (marginally) improved. Therefore, the findings with respect to predictable profits or losses for specific subgroups of the population will (likely) differ between risk equalization models for different years. Although most improvements were marginal and are not likely to have an important impact on the main conclusions of this dissertation, in 2024 a major change was implemented that has not been taken into account, since the most recent model replicated in this dissertation is that of 2022. With the update of the Dutch risk equalization model for 2024, several adjustments have been made not only to the risk adjusters but also to the method for estimating the payment weights of the model. This method, known as 'constrained regression', applies a restriction that ensures the predictable profits and losses after risk equalization for a specific definition[1] of (un)health are equal to zero. Due to the timing of the implementation of this significant adjustment to the model after the completion of the research within this dissertation, the quantitative results presented within this dissertation will differ from a scenario with constrained regression. Because the constraint applied in Dutch risk equalization is based on individual spending on pharmaceuticals in recent years (with such spending typical for chronically ill individuals), chronically ill individuals are – on average – better compensated by the model that incorporates constrained regression. Therefore, the predictable profits and losses for the subgroups of chronically ill individuals shown in this dissertation are expected to be larger than in a scenario where such a constraint is applied in the estimation of the payment weights. However, predictable profits and losses do remain after constrained regression, and therefore, so does the problem of risk selection. So, despite the deviation between the models replicated in this dissertation and the model currently in place in the Netherlands, the findings of this dissertation still present important implications for the design of risk equalization schemes.

7.3 DISCUSSING THE IMPLICATIONS OF THIS RESEARCH

The results of the five studies demonstrate that, despite structural improvements made over the years, sophisticated risk equalization leaves nontrivial incentives for risk selection for insurers. While all proposed strategies are successful in further reducing these incentives, when applied in isolation, they all involve tradeoffs with concerns over either the complexity of the model, incentives for cost control on the market, or affordability of health insurance coverage. Therefore, the benefits of each strategy must be carefully weighed against its costs.

1 The restriction is applied to the two classes of the former risk adjuster 'multiple-year pharmacy costs': individuals that have had pharmacy spending in the top 25% at least one of the past three years and those that have not. Therefore, through the constraint, both subgroups of the population have an average residual spending of zero.

In the next section, the findings of the studies are blended into an overarching discussion. Specifically, the aim is to discuss the broader implications for researchers and policymakers in terms of the evaluation of risk equalization, the design of risk equalization, and the supplementary strategies that could be used to reduce remaining selection incentives in the presence of risk equalization.

7.3.1 Improving the evaluation of risk equalization

One goal of this dissertation was to develop new evaluation measures for risk equalization performance to gain a better insight into the post-compensation incentives for risk selection for insurers. Typical benchmarks used in evaluation studies include basic statistical measures such as the R-squared and Cumming's Prediction Measure, which indicate the proportion of variance in medical spending that is ultimately explained by the payment system (Van Kleef et al., 2024b). Such measures provide insight into the raw functioning of the model but fail to provide a tangible evaluation of selection incentives (Van de Ven & Van Kleef, 2025). To quantify this crucial aspect within health insurance, the predictable profits and losses after risk equalization are a more suitable benchmark. Specifically, this result after risk equalization payments for subgroups of the population that are derived from data not explicitly included in the risk equalization model is notably relevant. Various studies have used such external data sources to quantify selection incentives, such as health survey information (Withagen-Koster et al., 2018), diagnoses of mental health disorders (Montz et al., 2016), or the use of specific prescription drugs (Geruso et al., 2019). However, such sources that help identify individuals with specific chronic diseases are not typically used in the structural evaluation of risk equalization. In the Netherlands, (modifications to) risk equalization is (are) evaluated through an evaluation framework. Recently, this framework has been updated in dialogue with researchers, insurers, and policymakers to incorporate meaningful metrics for risk equalization performance (WOR 1234). One new metric is the average financial result after risk equalization for two subgroups of individuals: those who are diagnosed with none of a list of 35 chronic diseases and those who are diagnosed with at least one. This is an important distinction with regard to health status and the overarching selection incentives towards these two subgroups. However, this dissertation demonstrates further improvements to risk equalization can still be made by expanding such measures within evaluation.

Although the distinction between no or any of the defined chronic diseases is valuable, a structural evaluation on the level of separate chronic diseases provides direct insight into the selection incentives for these specific groups. This is an important and necessary advancement to the evaluation of risk equalization, as chronically ill individuals are particularly vulnerable to actions of risk selection by insurers. For instance, quality-skimping by insurers may take place within the contracting of health services for specific, unprofitable disease groups (Van de Ven et al., 2015). Therefore, the availability of the Dutch general practitioner information from the Nivel-PCD as used within this dissertation presents an

outstanding opportunity to evaluate model performance for these vulnerable groups. The information from general practitioners facilitates a specific and objective identification of the health status of individuals, exceeding such potential of other available data sources. As shown in the research, the 109 identified groups of chronically ill individuals remain mostly undercompensated, despite the level of sophistication of the risk equalization model. So, incorporating (a range of) these 109 definitions of unhealth structurally in the evaluation of risk equalization should help to identify and address pressing needs for model development and risk selection threats.

Another aspect within the evaluation of risk equalization that could be improved is the uncertainty around the average financial result that is calculated for subgroups. The simulation in chapter three has shown that the variation in the financial result after risk equalization within and between subgroups (i.e., heteroscedasticity of residual spending) leads to selection incentives, even when the average result is zero. Given that insurers are typically risk averse, subgroups with less (more) uncertainty around the expected financial result might be relatively (un)attractive to insurers even if the average result of all subgroups is zero. Within the dissertation, two measures were applied to quantify the selection incentives, but both are theoretical applications and require further research before implementation into the evaluation framework. Also, the magnitude of this source of selection incentives is uncertain and deserves further attention. A starting point could be to investigate the behavior of insurers when facing heteroscedasticity in residual spending and their demonstrated degree of risk aversion. In the case that risk selection (incentives) related to heteroscedasticity in residual spending is indeed found, the variance in residual spending is an important addition to the evaluation (and improvement) of risk equalization. Combining the points discussed above, the evaluation of risk equalization can still be improved.

7.3.2 Improving risk equalization to further mitigate selection incentives

Over the past decades since risk equalization was first implemented, the models have gained substantial predictive accuracy, leading to a notable decline in selection incentives for competing insurers. However, as this dissertation has shown, nontrivial incentives for risk selection remain in the Dutch basic health insurance scheme. Therefore, the direct implication is that the prediction model has potential for further improvement to achieve better compensation for the wide range of risk types in the population. The logical follow-up is then to investigate how to achieve such further advancement. One direction traditionally followed in studies to improve risk equalization is the addition and modification of risk adjusters for the model. Within this dissertation, that route has been explored through an update to the Diagnosis-based Cost Groups of the Dutch model. Although the modification resulted in better compensation for specific subgroups of consumers, the subgroups of chronically ill individuals derived from the Nivel-PCD

remain mostly undercompensated. The adjustment therefore leads to better compensation for the individuals that are picked up by the risk adjuster itself but not necessarily for all subgroups of chronically ill consumers. So, to further mitigate selection incentives towards these vulnerable subgroups – and improve the functioning of the health system altogether – several modifications can be considered.

One example of a modification worth exploring is incorporating a multi-year feature for specific risk adjusters to better compensate for the (rise in) health spending related to (the development of) chronic diseases. In the current risk equalization design, qualifying for compensation through a risk adjuster hinges on the use of health services such as hospital treatments or pharmaceutical usage. Therefore, if such criteria are not met in a year where a patient with a chronic disease is less dependent on care (for instance, when the disease is developing slowly or intermittently), compensation to the insurer in the following year is (considerably) reduced. Instead, multi-year risk adjusters based on hospital treatments or pharmaceutical usage over multiple years (e.g., at least once in three years) would prevent such a drop in compensation. This could reduce selection incentives for insurers and yield additional benefits such as stimulating efficiency and prevention.

A next suggestion for improvement relates to the sources of information that can be used to derive risk adjusters. When new information sources that hold predictive value of future health spending become available, they should be considered for uptake in the model. For instance, information from other domains of health care, such as long-term care, could have predictive value for overall health spending. Moreover, such information is likely available for the entire population of insured individuals, which is an important requirement for uptake in the model. Further research should focus on exploring additional data sources that could improve risk equalization, are available for the entire insured population, and do not lead to undesired effects (e.g., higher compensation resulting from using information from providers with an inefficient practice style) (Stam et al., 2010). While uptake of the Nivel-PCD information from general practitioners would significantly improve the compensation for the identified groups of chronically ill individuals, its limited representativeness prevents such a scenario for now. Through further research and/or efforts by policymakers, this information could potentially be acquired for the entire population but requires substantial streamlining efforts in classification and reporting by general practitioners. Therefore, the search for alternative data sources may be a more promising avenue, at least in the short term.

Another worthwhile route for potential improvement of the risk equalization model is based on the dynamic interaction between consumer and insurer behavior. This aspect has been extensively studied, typically finding a negative correlation between the expected health spending of consumers and their elasticity to switch insurers. Within that context, Bijlsma et al. (2014), for instance, model the theoretically optimal design of risk

equalization systems, considering risk selection, efficiency, and consumer welfare. They argue that, from an efficiency point of view, to nullify selection incentives, risk equalization needs to overcompensate high-risk types individuals and leverage the competition that arises because of the relation between risk type and tendency to switch insurers. Although this specific discussion on risk equalization design was not directly addressed within this dissertation, the important dynamic of consumer-insurer interaction was studied within the context of risk selection.

Specifically, in chapter two of this dissertation, it was demonstrated that individuals who develop a new chronic disease are structurally undercompensated in the year of diagnosis. This is due to the prospective nature of the morbidity adjusters in the model designed to identify the presence of a chronic illness based on health utilization in the preceding year. Crucially, if these individuals anticipate the onset of a chronic condition when choosing a health insurer *before the year of diagnosis*, exploiting their information surplus relative to the chosen insurer, the expected financial loss *in the year of diagnosis* may be considered problematic. Therefore, changing the timeliness of morbidity adjusters that identify specific chronic diseases where such behavior is prevalent will directly improve the compensation and the negative influence on risk selection. A switch to concurrent risk equalization for such problematic chronic diseases will base the compensation for insurers on the treatment of these diseases in the current year. However, as a consequence, incentives for insurers to stimulate the prevention of (the aggravation of) these diseases will decrease because they face less financial risk related to the (increase in) health spending of their enrollees. Therefore, the benefits of such a modification should be carefully weighed against the potential downsides. Further research is required to identify which chronic diseases should be targeted for a switch to concurrent risk equalization. One example where concurrent risk equalization is employed in the Netherlands is for pregnant women or women who have delivered a child. This subgroup is undercompensated by €4,000 to €5,000 and typically switches to a comprehensive insurance plan before the year of expected child delivery (Gupta Strategists, 2023). As a result, insurers are financially discouraged from investing in care services of high quality for this specific subgroup. Therefore, since 2023, the Dutch risk equalization has incorporated a new, concurrent risk adjuster to detect and compensate pregnancy and child delivery in year t. This demonstrates that concurrent risk equalization can be a useful strategy for specific health changes that can be anticipated by consumers.

The last suggestion for changing the risk equalization model addresses the uncertainty around the financial result for subgroups. As proposed in section 7.3.1, the magnitude of the risk selection incentives that stem from heteroscedasticity in residual spending deserves further investigation. If indeed considered problematic from a perspective of risk selection, the variation in residual spending within and between subgroups could be incorporated in risk equalization. Future research aimed at the variety of strategies that

could be employed to compensate insurers for the higher uncertainty as a result of heteroscedasticity should help to identify the potential (dis)advantages and broader implications. For instance, a surcharge to risk equalization payment weights could be included for groups with problematic uncertainty, or an ex-post scheme could be designed to reduce the financial risk to insurers. This aspect shares similarities with the administrative costs for insurers that are not included in risk equalization and similarly differ between risk types (Douven et al., 2022a). Both aspects demonstrate that even if advancements made to the model remove predictable profits and losses, some incentives for risk selection may remain.

7.3.3 Supplementary strategies to mitigate selection incentives can be complementary

Although the implementation of constrained regression as the estimation method for Dutch risk equalization has significantly improved the compensation of chronically ill individuals, the scenario of complete absence of predictable profits and losses remains a distant goal (Van Kleef & Van Vliet, 2025). Therefore, in addition to further improvements of the risk equalization model, as proposed in section 7.3.2, supplementary strategies to mitigate risk selection could be explored. Two such strategies have been studied in this dissertation. First, high-risk pooling in chapter five as a modality of risk-sharing, where the availability of administrative information for multiple years of risk equalization has been used to identify and compensate a specific group of high-risk individuals. This strategy proved highly successful in exploiting the available information to reduce selection incentives, without sacrificing many incentives for cost-control. Putting high-risk pooling into practice, however, requires careful considerations by regulators. For instance, important considerations are the size of the budget designated for compensation, the financing of the compensation, and which entity selects which individuals will be allocated to the pool. Further research on such design questions is required and should involve policymakers and insurers to find the most effective and desirable outcome.

The other supplementary strategy explored in this dissertation is allowing insurers to risk-rate their premiums to some extent (see chapter 6). Allowing insurers to incorporate risk-rating in their premium-setting removes regulation-induced incentives for risk selection, which is an important advantage over premium-rate restrictions. Within this study, the sophisticated risk equalization model that was simulated was able to compensate insurers for 81% of the possible variation in risk-rated premiums. However, since 19% of the variation in premiums is *not* compensated by the risk equalization model, this could result in problems of affordability of coverage. For instance, within the simulation, 0.25% of the population pays an annual premium above €10,000. In addition, although risk-rated premiums may create more individual-level awareness of health care costs and the benefits of prevention, they may not only threaten affordability of coverage but also decrease transparency on the market, hindering competition and reducing the related

incentives for efficiency. Therefore, these tradeoffs resulting from removing or relaxing premium-rate restrictions deserve further attention before implementing such a strategy.

Importantly, the two supplementary strategies to mitigate incentives for risk selection discussed were studied in isolation. A combination of various strategies could alleviate the important downsides of the individual strategies and amplify their respective strengths. Therefore, while there is no silver bullet that alleviates all tradeoffs, combining the strategies could be considered by policymakers if these strategies have a complementary effect on reducing the remaining incentives for risk selection after risk equalization. For instance, if premium-rate restrictions are to be partially relaxed to ensure all premiums are constrained within a defined bandwidth in order to maintain affordability, risk-sharing could be applied for premiums that fall outside the agreed bandwidth. Within the bandwidth, a limited number of premium levels could be allowed to protect transparency on the market. Concerns over affordability are reduced, but incentives for cost-control are diminished through risk-sharing. However, the risk-sharing only applies for a small part of the population. Crucially, incentives for risk selection could be substantially reduced by combining these complementary strategies.

Again, there is no silver bullet, and tradeoffs are inevitable, but investigating a combination of supplementary strategies with complementary effects can be worthwhile for policymakers and an important topic for further research. Regulated competition among health insurers is an important construct to achieve efficiency, especially in the time of an ageing population and pressure on the financing of the system. However, within that context, it is of great importance that the focus of all entities involved is directed towards the efficiency and quality of health care rather than towards risk selection.

REFERENCES

REFERENCES

Andriola, C., Ellis, R. P., Siracuse, J. J., Hoagland, A., Kuo, T. C., Hsu, H. E., Walkey, A., Lasser, K. E., & Ash, A. S. (2024). A Novel Machine Learning Algorithm for Creating Risk-Adjusted Payment Formulas. *JAMA Health Forum, 5*(4), e240625-e240625. https://doi.org/10.1001/jamahealthforum.2024.0625

Ash, A., Porell, F., Gruenberg, L., Sawitz, E., & Beiser, A. (1989). Adjusting Medicare Capitation Payments Using Prior Hospitalization Data. *Health Care Financing Review, 10*(4), 17.

Bajari, P., Dalton, C., Hong, H., & Khwaja, A. (2014). Moral Hazard, Adverse Selection, and Health Expenditures: a Semiparametric Analysis. *The RAND Journal of Economics*, *45*(4), 747-763. https://doi.org/10.1111/1756-2171.12069

Battaglia, M. P., Hoaglin, D. C., & Frankel, M. R. (2009). Practical Considerations in Raking Survey Data. *Survey practice, 2*(5). https://doi.org/10.29115/SP-2009-0019

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. https://doi.org/10.1016/j.jpubeco.2012.05.011

Bauhoff, S., Fischer, L., Göpffarth, D. & Wuppermann, A. C. (2017). Plan Responses to Diagnosis-Based Payment: Evidence from Germany's Morbidity-Based Risk Adjustment. *Journal of Health Economics, 56*, 397-413. https://doi.org/10.1016/j.jhealeco.2017.03.001

Bayraktar, E., Milevsky, M. A., Promislow, S. D., & Young, V. R. (2009). Valuation of Mortality Risk via the Instantaneous Sharpe Ratio: Applications to Life Annuities. *Journal of Economic Dynamics and Control*, *33*(3), 676-691. https://doi.org/10.1016/j.jedc.2008.09.004

Bergquist, S. L., Layton, T. J., McGuire, T. G., & Rose, S. (2019). Data Transformations to Improve the Performance of Health Plan Payment Methods. *Journal of Health Economics*, *66*, 195-207. https://doi.org/10.1016/j.jhealeco.2019.05.005

Bijlsma, M., Boone, J., & Zwart, G. (2014). Competition leverage: how the demand side affects optimal risk adjustment. *The RAND Journal of Economics, 45*(4), 792-815. https://doi.org/10.1111/1756-2171.12071

Bolhaar, J., Lindeboom, M., & Van der Klaauw, B. (2012). A Dynamic Analysis of the Demand for Health Insurance and Health Care. *European economic review, 56*(4), 669-690. https://doi.org/10.1016/j.euroecorev.2012.03.002

Boonen, L. H., Laske-Aldershof, T., & Schut, F. T. (2016). Switching Health Insurers: the Role of Price, Quality and Consumer Information Search. *The European Journal of Health Economics*, *17*(3), 339-353. https://doi.org/10.1007/s10198-015-0681-1

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(9), 1359-1374. https://doi.org/10.1007/s10198-019-01102-w

Breiman, L. (2001). Random Forests. Machine Learning, 45(1), 5-32. https://doi.org/10.1023/A:1010933404324

Breyer, F., Bundorf, M. K., & Pauly, M. V. (2011). Health Care Spending Risk, Health Insurance, and Payment to Health Plans. In M. V. Pauly, T. G. McGuire & P. P. Barros (Eds.), *Handbook of Health*

Economics, *Volume 2*. (pp. 691-762). Elsevier. https://doi.org/10.1016/B978-0-444-53592-4.00011-6

Brief, R. P., & Owen, J. (1969). A Note on Earnings Risk and the Coefficient of Variation. *The Journal of Finance*, *24*(5), 901-904. https://doi.org/10.1111/j.1540-6261.1969.tb01701.x

Buchner, F., Wasem, J., & Schillo, S. (2017). Regression Trees Identify Relevant Interactions: can this improve the Predictive Performance of Risk Adjustment? *Health Economics, 26*(1), 74-85. https://doi.org/10.1002/hec.3277

Carey, C. (2017a). Technological Change and Risk Adjustment: Benefit Design Incentives in Medicare Part D. *American Economic Journal: Economic Policy, 9*(1), 38-73. https://doi.org/10.1257/pol.20140171

Carey, C. (2017b). *A Time to Harvest: Evidence on Consumer Choice Frictions from a Payment Revision in Medicare Part D*. [Working Paper]. Cornell University.

Croes, R. R., Schut, F. T., & Varkevisser, M. (2025). Adverse Selection and Consumer Inertia: Empirical Evidence from The Dutch Health Insurance Market. *The European Journal of Health Economics*, *26*(4), 641-651. https://doi.org/10.1007/s10198-024-01725-8

Cunningham, P. J., & Kohn, L. (2000). Health Plan Switching: Choice or Circumstance? Data from the Community Tracking Study give a Glimpse of Who Among the Privately Insured are Likely to Switch Plans, and Why. *Health Affairs*, *19*(3), 158-164. https://doi.org/10.1377/hlthaff.19.3.158

Cutler, D. M., & Zeckhauser, R. J. (2000). The Anatomy of Health Insurance. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics, Volume 1*. (pp. 563-643). Elsevier. https://doi.org/10.1016/S1574-0064(00)80170-5

Decarolis, F. & Guglielmo, A. (2017). Insurers' Response to Selection Risk: Evidence from Medicare Enrollment Reforms. *Journal of Health Economics, 56*, 383-396. https://doi.org/10.1016/j.jhealeco.2017.02.007

Diamond, P. (1992). Organizing the Health Insurance Market. *Econometrica: Journal of the Econometric Society, 60*(6), 1233-1254. https://doi.org/10.2307/2951520

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. In J. Kittler & F. Roli (Eds.), *Multiple Classifier Systems*. (pp 1-15). Springer Berlin. https://doi.org/10.1007/3-540-45014-9

Douven, R., Remmerswaal, M., & Mosca, I. (2015). Unintended Effects of Reimbursement Schedules in Mental Health Care. *Journal of Health Economics, 42*, 139-150. https://doi.org/10.1016/j.jhealeco.2015.03.008

Douven, R., Kauer, L., Demme, S., Paolucci, F., van de Ven, W. P., Wasem, J., & Zhao, X. (2022a). Should Administrative Costs in Health Insurance be Included in the Risk-Equalization? An Analysis of Five Countries. *The European Journal of Health Economics, 23*(9), 1437-1453. https://doi.org/10.1007/s10198-022-01436-y

Douven, R., & Van de Ven, W. P. (2022b). Neem bedrijfskosten van zorgverzekeraar ook op in de risicoverevening. *Economisch Statistische Berichten, 107*, 513-515. https://esb-nu.eur.idm.oclc.org/neem-bedrijfskosten-van-zorgverzekeraar-ook-op-in-de-risicoverevening/

Eeckhoudt, L., & Schlesinger, H. (2013). Higher-Order Risk Attitudes. In G. Dionne (Ed.) *Handbook of Insurance*. (pp. 41-57). Springer, New York, NY. https://doi.org/10.1007/978-1-4614-0155-1_2

Eijkenaar, F., Van Vliet, R. C., & Van Kleef, R. C. (2018). Diagnosis-Based Cost Groups in the Dutch Risk-Equalization Model: Effects of Clustering Diagnoses and of Allowing Patients to be Classified into Multiple Risk-Classes. *Medical care*, *56*(1), 91-96. https://doi.org/10.1097/MLR.0000000000000828

Einav, L., & Finkelstein, A. (2011). Selection in Insurance Markets: Theory and Empirics in Pictures. *Journal of Economic perspectives, 25*(1), 115-138. https://doi.org/115-138. 10.1257/jep.25.1.115

Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P., & Cullen, M. R. (2013). Selection on Moral Hazard in Health Insurance. *American Economic Review, 103*(1), 178-219. https://doi.org/178-219. 10.1257/aer.103.1.178

Einav, L., Finkelstein, A., & Tebaldi, P. (2024). *Market Design in Regulated Health Insurance Markets: Risk Adjustment vs. Subsidies* (No. w32586). National Bureau of Economic Research. https://doi.org/10.3386/w32586

Ellis, R. P. (1989). Employee Choice of Health Insurance. *The Review of Economics and Statistics*, *71*(2), 215-223. https://doi.org/10.2307/1926966

Ellis, R. P., & Ash, A. (1995). Refinements to the Diagnostic Cost Group (DCG) Model. *Inquiry, 32*(4), 418-429.

Ellis, R. P., Martins, B., & Zhu, W. (2017). Health Care Demand Elasticities by Type of Service. *Journal of Health Economics*, *55*, 232-243. https://doi.org/10.1016/j.jhealeco.2017.07.007

Ellis, R.P., Martins, B., & Rose, S. (2018). Risk Adjustment for Health Plan Payment. In T. G. McGuire & R. C. Van Kleef (Eds.), *Risk Adjustment, Risk-sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*. (pp. 55-104). Academic Press, Elsevier Inc. https://doi.org/10.1016/C2016-0-01299-8

Enthoven, A. C. (1978). Consumer-Choice Health Plan: A National Health-Insurance Proposal Based on Regulated Competition in the Private Sector (Second of Two Parts). *New England Journal of Medicine, 298*, 709-720. https://doi.org/10.1056/NEJM197803302981304

Enthoven, A. C. (1988a). Managed Competition: An Agenda for Action. *Health Affairs, 7*(3), 25-47. https://doi.org/10.1377/hlthaff.7.3.25

Enthoven, A. C. (1988b). Managed Competition of Alternative Delivery Systems. *Journal of Health Politics, Policy and Law, 13*(2), 305-321. https://doi.org/10.1215/03616878-13-2-305

Enthoven, A.C., & Kronick, R. (1989). A Consumer-Choice Health Plan for the 1990s. *New England Journal of Medicine, 320,* 29-37. https://doi.org/10.1056/NEJM198901053200106

European Insurance and Occupational Pensions Authority (EIOPA). (2015). *Cost-of-Capital rate.* EIOPA. https://www.eiopa.europa.eu/rulebook/solvency-ii/article-2452_en

EUR-Lex. (2009*). Directive 2009/138/EC of the European Parliament and of the Council of 25 November 2009 on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II)*. EUR-Lex. https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX%3A32009L0138

EUR-Lex. (2015a). *Commission Delegated Regulation (EU) 2015/35 of 10 October 2014 supplementing Directive 2009/138/EC of the European Parliament and of the Council on the taking-up and pursuit of the business of Insurance and Reinsurance (Solvency II).* EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015R0035

EUR-Lex. (2015b). *Commission Implementing Regulation (EU) 2015/2013 of 11 November 2015 laying down implementing technical standards with regard to standard deviations in relation to health risk equalisation systems in accordance with Directive 2009/138/EC of the European Parliament and of the Council.* EUR-Lex. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32015R2013&from=EN

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. https://doi.org/10.1016/j.jhealeco.2016.01.007

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. https://doi.org/10.1257/pol.20170014

Geruso, M., & Layton, T. (2020). Upcoding: Evidence from Medicare on Squishy Risk Adjustment. *Journal of Political Economy, 128*(3), 984-1026. https://doi.org/10.1086/704756

Glazer, J., & McGuire, T. G. (2000). Optimal Risk Adjustment in Markets with Adverse Selection: An Application to Managed Care. *American Economic Review*, *90*(4), 1055-1071. https://doi.org/10.1257/aer.90.4.1055

Gupta Strategists (2023). *Identificeren van zwangerschappen in de risicoverevening* (WOR 1161, Research report prepared for the Dutch Ministry of Health, Welfare and Sports). https://www.rijksoverheid.nl/documenten/rapporten/2023/08/23/identificeren-van-zwangerschappen-in-de-risicoverevening

Han, T., & Lavetti, K. (2017). Does Part D Abet Advantageous Selection in Medicare Advantage? *Journal of Health Economics, 56*, 368-382. https://doi.org/10.1016/j.jhealeco.2017.06.007

Hawkins, D. M. (2004). The Problem of Overfitting. *Journal of Chemical Information and Computer Sciences, 44*(1), 1-12. https://doi.org/10.1021/ci0342472

Het Nederlands Huisartsen Genootschap (2024). *NHG-Tabel-78 ICPC-1 NL-Engels-versie 1-Inkijkexemplaar.* Nederlands Huisartsen Genootschap. https://referentiemodel.nhg.org/sites/default/files/public/NHG-Tabel%2078-ICPC-1%20NL-Engels-versie%201-Inkijkexemplaar.pdf

Hodges, C. W., Taylor, W. R. L., & Yoder, J. A. (1997). Stocks, Bonds, the Sharpe Ratio, and the Investment Horizon. *Financial Analysts Journal, 53*(6), 74-80. https://doi.org/10.2469/faj.v53.n6.2132

Irvin, J. A., Kondrich, A. A., Ko, M., Rajpurkar, P., Haghgoo, B., Landon, B. E., Phillips, R. L., Petterson, S., Ng, A. Y., & Basu, S. (2020). Incorporating Machine Learning and Social Determinants of Health Indicators into Prospective Risk Adjustment for Health Plan Payments. *BMC Public Health, 20*(1), 608. https://doi.org/10.1186/s12889-020-08735-0

Kahane, Y. (1979). The Theory of Insurance Risk Premiums — a Re-Examination in the Light of Recent Developments in Capital Market Theory. *ASTIN Bulletin: The Journal of the IAA*, *10*(2), 223-239. https://doi.org/10.1017/S051503610000653X

Lamers, L. M. (1998). Risk-Adjusted Capitation Payments: Developing a Diagnostic Cost Groups Classification for the Dutch Situation. *Health Policy, 45*(1), 15-32. https://doi.org/10.1016/S0168-8510(98)00031-1

Lamers, L. M. (2001). Health-Based Risk Adjustment: is Inpatient and Outpatient Diagnostic Information Sufficient?. *Inquiry*, 423-431.

Lange, R., Schiller, J., & Steinorth, P. (2017). Demand and Selection Effects in Supplemental Health Insurance in Germany. *The Geneva Papers on Risk and Insurance-Issues and Practice*, *42*, 5-30. https://doi.org/10.1057/s41288-016-0023-2

Lavetti, K., & Simon, K. (2018). Strategic Formulary Design in Medicare Part D Plans. *American Economic Journal: Economic Policy, 10*, 154-192. https://doi.org/10.1257/pol.20160248

Layton, T. J., McGuire, T. G., & Sinaiko, A. D. (2016). Risk Corridors and Reinsurance in Health Insurance Marketplaces: Insurance for Insurers. *American Journal of Health Economics, 2*(1), 66-95. https://doi.org/10.1162/AJHE_a_00034

Layton, T. J., Ellis, R. P., McGuire, T. G., & Van Kleef, R. C. (2018a). Evaluating the Performance of Health Plan Payment Systems. In T. G. McGuire & R. C. Van Kleef (Eds.), *Risk adjustment, Risk-sharing and Premium Regulation in Health Insurance Markets: Theory and Practice.* (pp. 133-167). Academic Press, Elsevier Inc. https://doi.org/10.016/C2016-0-01299-8

Layton, T. J., McGuire, T. G., & Van Kleef, R. C. (2018b). Deriving Risk Adjustment Payment Weights to Maximize Efficiency of Health Insurance Markets. *Journal of health economics, 61*, 93-110. https://doi.org/10.1016/j.jhealeco.2018.07.001

Layton, T. J., Montz, E., & Shepard, M. (2018c). Health Plan Payment in US Marketplaces: Regulated Competition with a Weak Mandate. In T. G. McGuire & R. C. Van Kleef (Eds.), *Risk Adjustment, Risk-sharing and Premium Regulation in Health Insurance Markets: Theory and Practice.* (pp. 491-522). Academic Press, Elsevier Inc. https://doi.org/10.1016/C2016-0-01299-8

McGuire, T. G., & Van Kleef, R. C. (2018a). Regulated Competition in Health Insurance Markets: Paradigms and Ongoing Issues. In T. G. McGuire & R. C. Van Kleef (Eds.), *Risk Adjustment, Risk-sharing and Premium Regulation in Health Insurance Markets: Theory and Practice* (pp. 3-20). Academic Press, Elsevier Inc. https://doi.org/10.1016/B978-0-12-811325-7.00001-4

McGuire, T. G., & Van Kleef, R. C. (2018b). Risk-sharing. In T. G. McGuire & R. C. Van Kleef (Eds.), *Risk Adjustment, Risk-sharing and Premium Regulation in Health Insurance Markets: Theory and Practice* (pp. 105-131). Academic Press, Elsevier Inc. https://doi.org/10.1016/B978-0-12-811325-7.00004-X

McGuire, T.G., Schillo, S., & Van Kleef, R.C. (2020). Reinsurance, Repayments, and Risk Adjustment in Individual Health Insurance: Germany, the Netherlands and the U.S. marketplaces. American Journal of Health Economics, 6(1), 139-168. https://doi.org/10.1086/706796

McGuire, T. G., Schillo, S., & Van Kleef, R. C. (2021a). Very High and Low Residual Spenders in Private Health Insurance Markets: Germany, the Netherlands and the US Marketplaces. *European Journal of Health Economics*, *22*, 35-50. https://doi.org/10.1007/s10198-020-01227-3

McGuire, T. G., Zink, A. L., & Rose, S. (2021b). Improving the Performance of Risk Adjustment Systems: Constrained Regressions, Reinsurance, and Variable Selection. *American Journal of Health Economics, 7*(4), 497-521. https://doi.org/10.1086/716199

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. https://doi.org/10.1377/hlthaff.2011.1344

Montz, E., Layton, T., Busch, A. B., Ellis, R. P., Rose, S., & McGuire, T. G. (2016). Risk-Adjustment Simulation: Plans may have Incentives to Distort Mental Health and Substance Use Coverage. *Health Affairs, 35*(6), 1022-1028. https://doi.org/10.1377/hlthaff.2015.1668

Nederlandse Zorgautoriteit (NZa) (2019). *Monitor Zorgverzekeringen 2019*. https://puc.overheid.nl/nza/doc/PUC_289640_22/1/

Nederlandse Zorgautoriteit (NZa) (2020). *Monitor Zorgverzekeringen 2020*. https://open.overheid.nl/repository/ronl-b02f279f-7ab6-433d-8c65-e34171125a5c/1/pdf/monitor-zorgverzekeringen-2020.Pdf

Newhouse, J.P. (1996). Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection. *Journal of Economic Literature, 34*(3), 1236-1263

Nielen, M. M., Spronk, I., Davids, R., Korevaar, J. C., Poos, R., Hoeymans, N., Opstelten, W., Van der Sande, M. A., Biermans, M. C., Schellevis, F. G., & Verheij, R. A. (2019). Estimating Morbidity Rates Based on Routine Electronic Health Records in Primary Care: Observational Study. *JMIR medical informatics, 7*(3), e11929. https://doi.org/10.2196/11929

Nivel (2019). *Nivel Primary Care Database, NZR-00319.033*.

Nivel (2025). *Nivel Zorgregistraties Eerste Lijn - databron ziekten en zorg in de eerste lijn*. https://www.nivel.nl/nl/panels-en-registraties/nivel-zorgregistraties-eerste-lijn

Nuscheler, R., & Knaus, T. (2005). Risk Selection in the German public Health Insurance System. *Health Economics, 14*(12), 1253–1271. https://doi.org/10.1002/hec.1005

Olivieri, A., & Pitacco, E. (2015). Managing a Portfolio of Risks. In A. Olivieri & E. Pitacco (Eds.), *Introduction to Insurance Mathematics.* (pp. 75-155). Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-21377-4_2

Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2023). Improving Diagnosis-Based Cost Groups in the Dutch Risk Equalization Model: The Effects of a New Clustering Method and Allowing for Multimorbidity. *International Journal of Health Economics and Management, 23*, 1-22. https://doi.org/10.1007/s10754-023-09345-0

Oskam, M., Van Kleef, R. C., & Douven, R. (2024). Heteroscedasticity of Residual Spending After Risk Equalization: a Potential Source of Selection Incentives in Health Insurance Markets With Premium Regulation. *The European Journal of Health Economics, 25*(3), 379-396. https://doi.org/10.1007/s10198-023-01592-9

Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2025). Supplementing Risk Adjustment with High-Risk Pooling Using Historical Data for Identifying the High Risks. *Journal of Risk and Insurance, 92*(1), 166-202. https://doi.org/10.1111/jori.12500

Pauly, M. V., & Langwell, K. M. (1983). Research on competition in the market for Health Services: Problems and Prospects. *Inquiry, 20*(2), 142-161.

Prinsze, F. J., & Van Vliet, R. C. (2007). Health-Based Risk Adjustment: Improving the Pharmacy-Based Cost Group Model by Adding Diagnostic Cost Groups. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing, 44*(4), 469-480. https://doi.org/10.5034/inquiryjrnl_44.4.469

Renker, F., Häckl, D., & Wuppermann, A. (2025). Exploration of Using Constrained Regression in Germany's Morbidity-Based Risk Adjustment. *The European Journal of Health Economics*, 1-19. https://doi.org/10.1007/s10198-025-01819-x

Robinson, J. C., Gardner, L. B., & Luft, H. S. (1993). Health Plan Switching in Anticipation of Increased Medical Care Utilization. *Medical Care, 31*(1), 43-51.

Rose, S. (2016). A Machine Learning Framework for Plan Payment Risk Adjustment. *Health Services Research, 51*(6), 2358-2374. https://doi.org/10.1111/1475-6773.12464

Rose, S., Bergquist, S. L., & Layton, T. J. (2017). Computational Health Economics for Identification of Unprofitable Health Care Enrollees. *Biostatistics, 18*(4), 682-694. https://doi.org/10.1093/biostatistics/kxx012

Rothschild, M., & Stiglitz, J. (1976). Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information. *Quarterly Journal of Economics, 90*, 629-649.

Roy, M. H., & Larocque, D. (2012). Robustness of Random Forests for Regression. *Journal of Nonparametric Statistics, 24*(4), 993-1006. https://doi.org/10.1080/10485252.2012.715161

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*(2), 141-147. https://doi.org/10.1016/j.healthpol.2016.01.003

Schindler, C. J., Berndt, B., & Häckl, D. (2025). Pharmacy Cost Groups for the German Morbidity-Based Risk Compensation Scheme. *The European Journal of Health Economics*, 1-13. https://doi.org/10.1007/s10198-025-01809-z

Sharpe, W.F. (1994). The Sharpe Ratio. *Journal of Portfolio Management, 21*(1), 49-58.

Shen, Y., & Ellis, R. P. (2002). How Profitable is Risk Selection? A Comparison of Four Risk Adjustment Models. *Health economics, 11*(2), 165-174. https://doi.org/10.1002/hec.661

Shepard, M. (2022). Hospital Network Competition and Adverse Selection: Evidence from the Massachusetts Health Insurance Exchange. *American Economic Review, 112*(2), 578-615. https://doi.org/10.1257/aer.20201453

Shrestha, A., Bergquist, S., Montz, E., & Rose, S. (2018). Mental Health Risk Adjustment with Clinical Categories and Machine Learning. *Health Services Research, 53*, 3189-3206. https://doi.org/10.1111/1475-6773.12818

Stam, P. J., Van Vliet, R. C., & Van de Ven, W. P. (2010). A Limited-Sample Benchmark Approach to Assess and Improve the Performance of Risk Equalization Models. *Journal of Health Economics, 29*(3), 426-437. https://doi.org/10.1016/j.jhealeco.2010.02.001

Stolper, K.C., Boonen, L.H., Schut, F.T., & Varkevisser, M. (2022). Do Health Insurers use Target Marketing as a Tool for Risk Selection? Evidence from the Netherlands. *Health Policy, 126*(2), 122-128. https://doi.org/10.1016/j.healthpol.2021.12.005

Rosenberg, B., & Guy, J. (1976). Prediction of Beta from Investment Fundamentals: Part One. *Financial Analysts Journal*, *32*(3), 60-72. https://doi.org/10.2469/faj.v32.n3.60

Vanhommerig, J. W., Verheij, R. A., Hek, K., Ramerman, L., Hooiveld, M., Veldhuijzen, N. J., & Overbeek, L. I. (2025). Data Resource Profile: Nivel Primary Care Database (Nivel-PCD), the Netherlands. *International Journal of Epidemiology, 54*(2), dyaf017. https://doi.org/10.1093/ije/dyaf017

Van Barneveld, E. M., Van Vliet, R. C., & Van de Ven, W. P. (1996). Mandatory High-Risk Pooling: An Approach to Reducing Incentives for Cream Skimming. *Inquiry, 33*(2), 133-143.

Van Barneveld, E. M., Lamers, L. M., Van Vliet, R. C., & Van de Ven, W. P. (1998). Mandatory Pooling as a Supplement to Risk-Adjusted Capitation Payments in a Competitive Health Insurance Market. Social science & medicine, 47(2), 223-232. https://doi.org/10.1016/S0277-9536(98)00056-2

Van Barneveld, E. M., Lamers, L. M., Van Vliet, R. C., & Van de Ven, W. P. (2000). Ignoring Small Predictable Profits and Losses: a New Approach for Measuring Incentives for Cream Skimming. *Health Care Management Science, 3*(2), 131-140. https://doi.org/10.1023/A:1019029004807

Van Barneveld, E. M., Lamers, L. M., Van Vliet, R. C., & Van de Ven, W. P. (2001). Risk-sharing as a Supplement to Imperfect Capitation: a Tradeoff Between Selection and Efficiency. *Journal of Health Economics, 20*(2), 147-168. https://doi.org/10.1016/S0167-6296(00)00077-1

Van Barneveld, E. M., Van Vliet, R. C., & Van de Ven, W. P. (2001). Risk-sharing Between Competing Health Plans and Sponsors. *Health affairs*, *20*(3), 253-262. https://doi.org/10.1377/hlthaff.20.3.253

Van Kleef, R. C., Van De Ven, W. P., & Van Vliet, R. C. (2013a). Risk Selection in a Regulated Health Insurance Market: a Review of the Concept, Possibilities and Effects. *Expert review of pharmacoeconomics & outcomes research*, *13*(6), 743-752. https://doi.org/10.1586/14737167.2013.841546

Van Kleef, R. C., Van Vliet, R. C., & Van de Ven, W. P. (2013b). Risk Equalization in the Netherlands: an Empirical Evaluation. *Expert review of pharmacoeconomics & outcomes research, 13*(6), 829-839. https://doi.org/10.1586/14737167.2013.842127

Van Kleef, R. C., Van Vliet, R. C., & Van Rooijen, E. M. (2014). Diagnoses-Based Cost Groups in the Dutch Risk-Equalization Model: the Effects of Including Outpatient Diagnoses. *Health policy, 115*(1), 52-59. https://doi.org/10.1016/j.healthpol.2013.07.005

Van Kleef, R. C., McGuire, T. G., Van Vliet, R. C., & Van de Ven, W. P. (2017). Improving Risk Equalization with Constrained Regression. *The European Journal of Health Economics*, *18*(9), 1137-1156. https://doi.org/10.1007/s10198-016-0859-1

Van Kleef, R. C., Eijkenaar, F., Van Vliet, R. C., & Van de Ven, W. P. (2018a). Health Plan Payment in the Netherlands. In T. G. McGuire & R. C. Van Kleef (Eds.), *Risk Adjustment, Risk-sharing and Premium Regulation in Health Insurance Markets: Theory and Practice* (pp. 397-429). Academic Press, Elsevier Inc. https://doi.org/10.1016/B978-0-12-811325-7.00014-2

Van Kleef, R. C., Schut, F. T., & Van de Ven, W. P. (2018b). Premium Regulation, Risk Equalization, Risk-sharing, and Subsidies: Effects on Affordability and Efficiency. In T. G. McGuire & R. C. Van Kleef (Eds.), *Risk Adjustment, Risk-sharing and Premium Regulation in Health Insurance Markets: Theory and Practice.* (pp. 21-54). Academic Press, Elsevier Inc. https://doi.org/10.1016/C2016-0-01299-8

Van Kleef, R. C., Eijkenaar, F., & Van Vliet, R. C. (2019). Selection Incentives for Health Insurers in the Presence of Sophisticated Risk Adjustment. *Medical Care Research and Review, 77*(6), 584-595. https://doi.org/10.1177/1077558719825982

Van Kleef, R. C., Eijkenaar, F., van Vliet, R. C., & Nielen, M. M. (2020a). Exploiting Incomplete Information in Risk Adjustment Using Constrained Regression. *American Journal of Health Economics, 6*(4), 477-497. https://doi.org/10.1086/710526

Van Kleef, R. C., Oskam, M., Visser, L. A., Van Vliet, R. C., & Withagen-Koster, A. A. (2020b). Grootonderhoud DKG's t.b.v. somatisch model 2021 (WOR 988, Research report prepared for the Dutch Ministry of Health, Welfare and Sports). Rotterdam: Erasmus University Rotterdam https://open-overheid-nl.eur.idm.oclc.org/repository/ronl-541452cb-600a-4ada-9847-682de4317410/1/pdf/WOR%20988%20-%20Grootonderhoud%20DKGs%20tbv%20somatisch%20model%2020b21.pdf

Van Kleef, R. C., Reuser, M., Stam, P. J., & Van de Ven, W. P. (2022a). *Positive and Negative Effects of Risk Equalization and Risk-sharing in regulated Competitive Health Insurance Markets*. [EsCHER Working Paper No. 2022014]. Erasmus University Rotterdam. https://www.eur.nl/en/media/2022-11-2022014-kleef-et-al

Van Kleef, R. C., & Van Vliet, R. C. (2022b). How to Deal with Persistently Low/High Spenders in Health Plan Payment Systems? *Health Economics, 31*(5), 784-805. https://doi.org/10.1002/hec.4477

Van Kleef, R. C., Reuser, M., McGuire, T. G., Armstrong, J., Beck, K., Brammli-Greenberg, S., & Wasem, J. (2024a). Scope and Incentives for Risk Selection in Health Insurance Markets With Regulated Competition? An International Comparison. *Medical Care Research and Review, 81*(3), 175-194. https://doi.org/10.1177/10775587231222584

Van Kleef, R. C., Reuser, M., Stam, P. J., & Van de Ven, W. P. (2024b). A Framework for Ex-Ante Evaluation of the Potential Effects of Risk Equalization and Risk-sharing in Health Insurance Markets with Regulated Competition. *Health Economics Review*, *14*(1), 57. https://doi.org/10.1186/s13561-024-00540-4

Van Kleef, R. C., & Van Vliet, R. C. (2025). Evaluatie Risicoverevening 2024 & 2025. https://www.eur.nl/en/eshpm/media/2025-03-eindrapport-evaluatie-risicoverevening-2024-2025

Van Veen, S. H., Van Kleef, R. C., Van de Ven, W. P., & Van Vliet, R. C. (2015). Is There One Measure-Of-Fit That Fits All? A Taxonomy and Review of Measures-Of-Fit For Risk-Equalization Models. *Medical Care Research and Review, 72*(2), 220-243. https://doi.org/10.1177/1077558715572900

Van Veen, S. H., Van Kleef, R. C., Van de Ven, W. P., & Van Vliet, R. C. (2018). Exploring the Predictive Power of Interaction Terms in a Sophisticated Risk Equalization Model Using Regression Trees. *Health Economics, 27*(2), e1-e12. https://doi.org/10.1002/hec.3523

Van de Ven, W. P., & Van Praag, B. M. (1981). The Demand for Deductibles in Private Health Insurance: A Probit Model with Sample Selection. *Journal of Econometrics, 17*(2), 229-252. https://doi.org/10.1016/0304-4076(81)90028-2

Van de Ven, W. P., & Ellis, R. P. (2000a). Risk Adjustment in Competitive Health Plan Markets. In A. J. Culyer & J. P. Newhouse (Eds.), *Handbook of Health Economics, Volume 1*. (pp. 755-845). Elsevier Inc., Amsterdam. https://doi.org/10.1016/S1574-0064(00)80173-0

Van de Ven, W. P., Van Vliet, R. C., Schut, F. T., & Van Barneveld, E. M. (2000b). Access to Coverage for High-Risks in a Competitive Individual Health Insurance Market: Via Premium Rate Restrictions or Risk-Adjusted Premium Subsidies?. *Journal of Health Economics, 19*(3), 311-339. https://doi.org/10.1016/S0167-6296(99)00028-4

Van de Ven, W. P., & Schut, F. T. (2011). Guaranteed Access to Affordable Coverage in Individual Health Insurance Markets. In S. Glied & P. C. Smith (Eds.), *The Oxford Handbook of Health Economics.* (pp. 380-404). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199238828.013.0017

Van de Ven, W. P., Van Kleef, R. C., Van Vliet, R. C. (2015). Risk Selection Threatens Quality of Care for Certain Patients: Lessons From Europe's Health Insurance Exchanges. *Health Affairs, 34*(10), 1713-1720. https://doi.org/10.1377/hlthaff.2014.1456

Van de Ven, W. P., Van Vliet, R. C., & 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*(2), 167–180. https://doi.org/10.1007/s10198-016-0764-7

Van de Ven, W. P., & Van Kleef, R. C. (2025). A Critical Review of the Use of R^2 in Risk Equalization Research. *The European Journal of Health Economics, 26*(3), 363-375. https://doi.org/10.1007/s10198-024-01709-8

Van Vliet, R. C. (2006). Free Choice of Health Plan Combined With Risk-Adjusted Capitation Payments: Are Switchers and New Enrolees Good Risks? *Health Economics, 15*(8), 763-774. https://doi.org/10.1002/hec.1102

Van Vliet, R. C, Van Kleef, R. C., & Van Rooijen, E. M. (2015). *Groot onderhoud van de diagnosekostengroepen (DKG's) in het risicovereveningsmodel voor de somatische zorgkosten* (WOR 728 Research report prepared for the Dutch Ministry of Health, Welfare and Sports). Rotterdam: Erasmus University Rotterdam.

Vimont, A., Leleu, H., & Durand-Zaleski, I. (2022). Machine Learning Versus Regression Modelling in Predicting Individual Healthcare Costs From a Representative Sample of the Nationwide Claims Database in France. *The European Journal of Health Economics, 23*(2), 211-223. https://doi.org/10.1007/s10198-021-01363-4

Wang, C. D., Chen, Z., Lian, Y., & Chen, M. (2022). Asset Selection Based on High Frequency Sharpe Ratio. *Journal of Econometrics*, *227*(1), 168-188. https://doi.org/10.1016/j.jeconom.2020.05.007

Whitbeck, V. S., & Kisor Jr, M. (1963). A New Tool in Investment Decision-Making. *Financial Analysts Journal*, *19*(3), 55-62. https://doi.org/10.2469/faj.v19.n3.55

Withagen-Koster, A. A., Van Kleef, R. C., & Eijkenaar, F. (2018). Examining unpriced Risk Heterogeneity in the Dutch Health Insurance Market. *The European Journal of Health Economics, 19*(9), 1351-1363. 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. *The European Journal of Health Economics, 21*(4), 513. https://doi.org/10.1007/s10198-019-01146-y

Withagen-Koster, A. A. (2022). *Evaluating and Improving Risk Equalization with Health Survey Information* [Doctoral Thesis, Erasmus University Rotterdam]. https://pure.eur.nl/ws/portalfiles/portal/68792002/anja_withagen_koster_thesis_63500e1b85897.pdf

Withagen-Koster, A. A., van Kleef, R. C., & Eijkenaar, F. (2023). Predictable Profits and Losses in a Health Insurance Market With Risk Equalization: A Multiple-Contract Period Perspective. *Health Policy, 131*, 104763. https://doi.org/10.1016/j.healthpol.2023.104763

Withagen-Koster, A. A., Van Kleef, R. C., & Eijkenaar, F. (2024). High-Risk Pooling for Mitigating Risk Selection Incentives in Health Insurance Markets With Sophisticated Risk Equalization: an Application Based on Health Survey Information. *BMC Health Services Research*, *24*(1), 273. https://doi.org/10.1186/s12913-024-10774-x

WOR 1234. WOR-Toetsingskader Risicoverevening. Den Haag: Ministerie van VWS.

Wright, M. N., Wager, S., & Probst, P. (2019). A Fast Implementation of Random Forests. *R package version 0.11*, *2*, 123-136. https://mran.revolutionanalytics.com/snapshot/2016-01-01/web/packages/ranger/ranger.pdf

Xu, J., Trish, E., & Joyce, G. (2021). Incorporating Prescription Drug Utilization Information Into the Marketplace Risk Adjustment Model Improves Payment Accuracy and Reduces Adverse Selection Incentives. *Medical Care Research and Review, 78*(4), 381-391. https://doi.org/10.1177/1077558719870060

Xu, W., Cai, G. J., Li, G. N., Cao, J. J., Shi, Q. H., & Bai, J. (2016). Age or Health Status: Which Influences Medical Insurance Enrollment Greater? *Journal of Global Health, 6*(2), 020801. https://doi.org/10.7189/jogh.06.020801

Zink, A., & Rose, S. (2021) Identifying Undercompensated Groups Defined by Multiple Attributes in Risk Adjustment. *BMJ Health & Care Informatics, 28*(1), e100414. https://doi.org/10.1136/bmjhci-2021-100414

Zorginstituut Nederland (2020). *Regeling Risicoverevening 2020*. https://www.zorginstituutnederland.nl/site/binaries/site-content/collections/documents/2019/10/02/zvw-2020/Regeling_risicoverevening_2020_%28staatscourant%29.pdf

Zorginstituut Nederland (2026). *Risicoverevening Zvw*. https://www.zorginstituutnederland.nl/financiering/informatie-voor-zorginstanties-verzekeraars-en-zorgkantoren/risicoverevening-zvww

GENERAL APPENDIX

Translation of ICPC list		
Code	Name of chronic condition in English	Name of chronic condition in Dutch
General and Unspecified		
A28	Limited function/disability, not otherwise specified	Functiebeperking/handicap
A79	Malignancy, not otherwise specified	Maligniteit met onbekende primaire lokalisatie
A90	Congenital anomaly, not otherwise specified/multiple	Multiple aangeboren afwijkingen
Blood, Blood Forming Organs and Immune Mechanism		
B28	Limited function/disability	Functiebeperking/handicap bloed/lymfestelse
B72	Hodgkin's disease/lymphoma	Ziekte van Hodgkin
B73	Leukaemia	Leukemie
B74	Malignant neoplasm blood, other	Andere maligniteit bloed/lymfestelsel
B78	Hereditary haemolytic anaemia	Erfelijke hemolytische anemie
B79	Congenital anomaly blood/lymph, other	Andere aangeboren afwijking bloed/lymfestel
B83	Purpura/coagulation defect	Purpura/stollingsstoornis/afwijkende trombocyten
B90	HIV-infection/aids	HIV-infectie (AIDS/ARC)
Digestive system		
D28	Limited function/disability	Functiebeperking/handicap spijsverteringsorganen
D74	Malignant neoplasm stomach	Maligniteit maag
D75	Malignant neoplasm colon/rectum	Maligniteit colon/rectum
D76	Malignant neoplasm pancreas	Maligniteit pancreas
D77	Malignant neoplasm digestive, other/not otherwise specified	Andere/niet-gespecificeerde maligniteit spijsvertering-sorganen
D81	Congenital anomaly digestive system	Aangeboren afwijking(en) spijsverteringsorganen
D92	Diverticular disease intestines	Diverticulose/diverticulitis
D94	Chronic enteritis/ulcerative colitis	Colitis ulcerosa/chronische enteritis (regionalis)
D97	Cirrhosis/Liver disease, not otherwise specified	Cirrose/andere leverziekte
Eye		
F28	Limited function/disability	Functiebeperking/handicap oog/adnexen
F81	Congenital anomaly eye, other	Andere aangeboren afwijking(en) oog/adnexen
F83	Retinopathy	Retinopathie
F84	Macular degeneration	Maculadegeneratie
F91	Refractive error	Refractie afwijking(en)
F93	Glaucoma	Glaucoom/verhoogde oogdruk
F94	Blindness	Blindheid (elke graad/vorm)
Ear		
H28	Limited function/disability	Functiebeperking/handicap oor
H80	Congenital anomaly of ear	Aangeboren afwijking(en) oor
H83	Otosclerosis	Otosclerose
H84	Presbycusis	Presbyacusis
H85	Acoustic trauma	Akoestisch letsel/lawaaidoofheid

Translation of ICPC list		
Code	Name of chronic condition in English	Name of chronic condition in Dutch
H86	Deafness	Doofheid/slechthorendheid
Cardiovascular		
K28	Limited function/disability	Functiebeperking/handicap hartvaatstelsel
K73	Congenital anomaly cardiovascular	Aangeboren afwijking(en) hartvaatstelsel
K74	Angina pectoris	Angina pectoris
K76	Other and chronic ischaemic heart disease	Andere/chronische ischemische hartziekte
K77	Heart failure	Decompensatio cordis
K82	Pulmonary heart disease	Cor pulmonale
K86	Hypertension uncomplicated	Essentiële hypertensie zonder orgaanbeschadiging
K87	Hypertension complicated	Hypertensie met orgaanbeschadiging/secundaire hypertensie
K90	Stroke/cerebrovascular accident	Cerebrovasculair accident (CVA)
K91	Atherosclerosis	Atherosclerose
K92	Other arterial obstruction / peripheral vascular disease	Andere ziekte(n) perifere arteriën
Musculoskeletal		
L28	Limited function/disability	Functiebeperking/handicap bewegingsapparaat
L82	Congenital anomaly musculoskeletal	Aangeboren afwijking(en) bewegingsapparaat
L84	Osteoarthritis of spine (any region)	Artrose/spondylose wervelkolom
L85	Acquired deformity of spine	Verworven afwijking(en) wervelkolom
L88	Rheumatoid arthritis and allied conditions	Reumatoïde artritis/verwante aandoening(en)
L89	Osteoarthrosis of hip	Coxartrose
L90	Osteoarthrosis of knee	Gonartrose
L91	Osteoarthrosis, other	Andere artrose/verwante aandoening(en)
L95	Osteoporosis	Osteoporose
L98	Acquired deformity of limb	Verworven afwijking(en) extremiteiten
Neurological		
N28	Limited function/disability	Functiebeperking/handicap zenuwstelsel
N70	Poliomyelitis/other enterovirus	Poliomyelitis/andere enterovirus infectie
N74	Malignant neoplasm nervous system	Maligniteit zenuwstelsel
N85	Congenital anomaly neurological	Aangeboren afwijking(en) zenuwstelsel
N86	Multiple sclerosis	Multiple sclerose
N87	Parkinsonism/paralysis agitans	Parkinsonisme, ziekte van Parkinson
N88	Epilepsy, all types	Epilepsie (alle vormen)
Psychological		
P28	Limited function/disability	Functiebeperking/handicap psychische ziekte
P70	Dementia	Dementie
P72	Schizophrenia, all types	Schizofrenie
P80	Personality disorder	Persoonlijkheids-/karakterstoornis
P85	Mental retardation	Mentale retardatie/intellectuele achterstand

Translation of ICPC list		
Code	Name of chronic condition in English	Name of chronic condition in Dutch
Respiratory		
R28	Limited function/disability	Functiebeperking/handicap luchtwegen
R84	Malignant neoplasm trachea/bronchus/lung	Maligniteit bronchus/long
R85	Malignant neoplasm respiratory, other	Andere maligniteit luchtwegen
R89	Congenital anomaly respiratory	Aangeboren afwijking(en) luchtwegen
R91	Chronic bronchitis/ bronchiectasis	Chronische bronchitis/bronchiëctasieën
R95	Emphysema/chronic obstructive pulmonary disease	Emfyseem/COPD
R96	Asthma	Astma
Skin		
S28	Limited function/disability	Functiebeperking/handicap huid/subcutis
S77	Malignant neoplasm of skin	Maligniteit huid/subcutis
S81	Haemangioma/lymphangioma	Hemangioom/lymfangioom
S83	Congenital skin anomaly, other	Andere aangeboren afwijking(en) huid/subcutis
S87	Atopic dermatitis/eczema	Constitutioneel eczeem
S91	Psoriasis	Psoriasis
Endocrine/Metabolic and Nutritional		
T28	Limited function/disability	Functiebeperking/handicap endocriene klieren/metabolisme/voeding
T71	Malignant neoplasm thyroid	Maligniteit schildklier
T78	Thyroglossal duct cysts	Persisterende ductus thyreoglossus/cyste
T80	Congenital anomaly endocrine/ metabolic	Andere aangeboren afwijking endocriene klieren/metabolisme
T81	Goitre, thyroid nodule including thyrotoxicosis	Struma/noduli
T86	Hypothyroidism/myxoedema	Hypothyreoïdie/myxoedeem
T90	Diabetes mellitus	Diabetes mellitus
T92	Gout	Jicht
T93	Lipid metabolism disorder	Vetstofwisselingsstoornis(sen)
Urological		
U28	Limited function/disability urinary	Functiebeperking/handicap urinewegen
U75	Malignant neoplasm of kidney	Maligniteit nier
U76	Malignant neoplasm of bladder	Maligniteit blaas
U77	Malignant neoplasm urinary, other	Andere maligniteit urinewegen
U85	Congenital anomaly urinary tract	Aangeboren afwijking(en) urinewegen
U88	Glomerulonephritis/nephrosis	Glomerulonephritis/nefrose
Pregnancy, Childbearing, Family Planning		
W28	Limited function/disability	Functiebeperking/handicap ten gevolge van zwangerschap
W72	Malignant neoplasm coexisting with pregnancy	Maligniteit in verband met zwangerschap

Translation of ICPC list		
Code	Name of chronic condition in English	Name of chronic condition in Dutch
W76	Congenital anomaly complicating pregnancy	Zwangerschap complicerende aangeboren afwijking moeder
Female Genital		
X28	Limited function/disability	Functiebeperking/handicap geslachtsorganen vrouw
X75	Malignant neoplasm cervix	Maligniteit cervix uteri
X76	Malignant neoplasm breast female	Maligniteit borst vrouw
X77	Malignant neoplasm female genital, other	Andere maligniteit geslachtsorganen vrouw
X83	Congenital anomaly genital female	Aangeboren afwijking(en) geslachtsorganen vrouw
X88	Chronic cystic disease breast female	Fibroadenoom/polycystische afwijking borsten
Male Genital		
Y28	Limited function/disability	Functiebeperking/handicap geslachtsorganen man
Y77	Malignant neoplasm prostate	Maligniteit prostaat
Y78	Malignant neoplasm male genital, other	Andere maligniteit geslachtsorganen/borsten man
Y82	Hypospadias	Hypospadie
Y84	Congenital male genital anomaly, other	Andere aangeboren afwijking(en) geslachtsorganen/ borsten man
Social Problems		
Z28	Limited function/disability	Sociale functiebeperking/handicap

SUMMARY

SUMMARY

Introduction

A health insurance system can play a significant role in ensuring the financial accessibility of health care. However, without regulation, premiums are adjusted based on the expected health care costs of individuals, which means that those with higher health risks, in particular, face high or even unaffordable premiums. To ensure accessibility and solidarity within the health care system, several countries – including the Netherlands, Germany, Switzerland, and the United States – have established systems that combine competition among health insurers with extensive government regulation. Key elements of this regulation include an obligation to accept all applicants, a standardized basic coverage package, and restrictions on premium differentiation.

Although this combination of regulations safeguards the affordability and accessibility of health insurance, it simultaneously creates financial incentives for health insurers to attract individuals with low health risks and, conversely, to avoid those with high health risks. After all, when all policyholders pay the same premium but generate predictably different health care costs, incentives for risk selection by health insurers arise. They can then act on these incentives in the design of insurance products and marketing strategies or, for example, through selective contracting with health care providers. Risk selection can thus put pressure on the accessibility, quality, and efficiency of health care and distort fair competition among health insurers.

To limit these incentives, various systems with regulated competition employ the instrument of risk equalization. The risk equalization system aims to provide health insurers with risk-based compensation for the predictable differences in health care costs among their insured individuals. As the risk equalization model becomes better able to predict differences in expected health care costs, the predictable profits and losses on specific groups of insured individuals decrease, thereby also reducing the incentives for risk selection. Over the past few decades, the models used have therefore evolved from relatively simple models based on demographic characteristics to advanced models that utilize in-depth information on health status and health care utilization.

Despite these ongoing improvements, both national and international studies show that even the most advanced risk equalization models still result in substantial undercompensation and overcompensation for specific groups of insured individuals, particularly those with chronic conditions. As a result, financial incentives for risk selection persist. Further improvement of risk equalization therefore remains necessary to safeguard the accessibility, solidarity, and efficiency of health insurance markets.

Reducing selection incentives requires, on the one hand, a careful evaluation of the extent to which risk equalization actually succeeds in compensating for predictable cost differences among insured individuals. On the other hand, it is necessary to develop and test additional strategies that can further reduce the remaining selection incentives. This dissertation contributes to both objectives by developing new methods for evaluating risk equalization and by examining various policy options that can further improve the functioning of regulated health insurance markets.

Research Objective

The objective of this dissertation is therefore to develop new methods to better evaluate the selection incentives that persist after risk equalization and to investigate strategies to further reduce these selection incentives. This dissertation aims to contribute to the further development of risk equalization by both introducing new evaluation methods and developing and empirically testing various policy strategies.

The first part of the dissertation focuses on evaluating the effectiveness of risk equalization using innovative methods. To this end, new approaches are developed that utilize objective general practitioner data and administrative data to assess the effectiveness of risk equalization from new perspectives. The aim of the first part of the dissertation is thus to increase understanding of the nature, extent, and dynamics of residual selection incentives.

The second part of the dissertation focuses on simulating strategies that can further reduce these selection incentives. The first strategy involves improving an existing feature of the Dutch risk equalization model. The second strategy explores the possibilities of an innovative form of risk sharing as a supplement to risk equalization. The third strategy explores the extent to which a partial relaxation of premium regulation can contribute to reducing selection incentives. By analyzing these three complementary strategies, insight is gained into the trade-offs policymakers face regarding limiting risk selection, maintaining efficiency incentives, ensuring solidarity, and promoting access to health care.

Part I: Evaluating Risk Equalization Through Innovative Methods

Although risk equalization models are typically assessed based on the predictable profits and losses for specific groups of insured individuals, common evaluation methods have various limitations. For example, these methods often rely on a single year of observation, and selection incentives are assessed solely on the basis of average undercompensation and overcompensation for defined groups of insured individuals. As a result, important aspects of selection incentives are overlooked. The first part of this dissertation therefore introduces two new approaches that broaden the evaluation of risk equalization. These approaches utilize objective general practitioner data from the Nivel Primary Care Database and administrative data from the Dutch risk equalization system.

Chapter 2 examines the extent to which prospective risk equalization adequately compensates health insurers for insured individuals diagnosed with a new chronic condition. Previous evaluation studies often treat individuals with a chronic condition as a single homogeneous group, whereas the timing of the onset of a condition is expected to have a significant impact on related health care costs and the extent to which health insurers are compensated. Because prospective risk equalization relies on information from previous years, a sudden deterioration in health cannot be immediately reflected in the compensation that health insurers receive.

Using data from general practitioners, individuals are identified who are diagnosed with a chronic condition for the first time in a given year. Their health care spending and risk equalization payments are then tracked from two years prior to through two years after the diagnosis. In addition, the study examines the extent to which these insured individuals switch health insurers around the year of diagnosis, as such switches can influence risk selection by health insurers. The results show that health insurers face substantial losses in the year a chronic condition is first diagnosed. Significant undercompensation also persists for various conditions in the years both before and after diagnosis. At the same time, it appears that insured individuals respond to changes in their health status when choosing a health insurer. The combination of these findings suggests that undercompensation may lead to significant selection incentives for health insurers. The study thus demonstrates that taking into account the dynamics surrounding the onset of chronic conditions provides new insights into selection incentives and that further improvement of the risk equalization model remains necessary.

Chapter 3 then introduces a second innovative approach to evaluating risk equalization. Whereas existing evaluations focus almost exclusively on the average predictable profits and losses of insured individuals, this chapter examines another source of selection incentives: the distribution of remaining health care costs after contributions from risk equalization. Even when health insurers are, on average, fully compensated for the health care costs of different groups of insured individuals, significant differences may exist in the uncertainty surrounding the insurers' final financial results. This chapter demonstrates that the residual health care costs after risk equalization are heteroscedastic: groups with higher expected health care costs typically also exhibit greater variation in the remaining financial results for health insurers. For risk-averse health insurers, this greater uncertainty means that enrolling such insured individuals may be less attractive, even when the average financial outcome is neutral. In addition, it leads to higher capital requirements under the solvency supervision of health insurers. Through these mechanisms, heteroscedasticity of residual spending can constitute an independent source of selection incentives that has largely been overlooked in the existing literature.

The findings of Chapters 2 and 3 collectively show that selection incentives are not fully captured by looking exclusively at the average undercompensation and overcompensations for statically defined groups of insured individuals. Both the evolution of insured individuals' health status over time and the uncertainty surrounding financial outcomes after risk equalization prove to be additional dimensions relevant to evaluating the effectiveness of risk equalization. Thus, the first part of this dissertation broadens the toolkit for assessing the effectiveness of risk equalization systems and provides a more complete picture of the selection incentives that persist despite advanced risk equalization.

Part II: Simulating Strategies to Reduce Selection Incentives

Whereas the first part of this dissertation focuses on better understanding and evaluating the remaining selection incentives after risk equalization, the second part centers on the question of how these selection incentives can be further reduced. This part of the dissertation therefore examines three complementary policy strategies to further limit these incentives. The first strategy focuses on improving the risk equalization model itself, the second on supplementing risk equalization with an innovative form of risk sharing, and the third on reevaluating the regulations regarding premium differentiation.

Chapter 4 examines the extent to which selection incentives can be reduced by further improving one of the health-related equalization characteristics in the Dutch risk equalization model. The "diagnosis-based cost groups" morbidity adjuster is designed to identify the presence of chronic conditions based on hospital claims and to compensate health insurers accordingly. The goal of modifying this adjuster is to improve the accuracy of the equalization model by applying a new clustering technique to the underlying diagnoses and treatments and by allowing multiple classifications of insured individuals within the adjuster. To objectively evaluate this improvement, general practitioner data from the Nivel Primary Care Database are used to determine the presence of a chronic condition and then to assess how the model improvement affects the distinct groups.

The results show that the model adjustment leads to a significant reduction in predictable profits and losses for groups of insured individuals identified using claims data and for those with multimorbidity. However, for the groups of insured individuals identified using general practitioner data, the results are less convincing. This demonstrates that, despite the model improvement, insured individuals with a chronic condition remain largely undercompensated by the risk equalization model. Further improvement of the model is therefore necessary to enhance efficiency and accessibility for these vulnerable groups and the health care system as a whole.

Chapter 5 then examines a distinctly different approach. Instead of further expanding the risk equalization model, an innovative form of risk sharing – high-risk pooling – is simulated as a supplement to risk equalization. Under the high-risk pooling system, pre-

selected insured individuals are assigned to a separate pool if they are expected to have exceptionally high remaining health care spending after the risk equalization payments have been allocated. The insurers of these individuals are then compensated through a targeted contribution for the final mismatch between their health care spending and the equalization contribution. In this study, various strategies are applied to historical data to identify a group of high-risk insured individuals. Next, the size of the pool was determined to specifically reduce selection incentives while preserving efficiency incentives as much as possible.

In this chapter, various variants of high-risk pooling and alternative forms of risk sharing are developed and compared. The analyses show that a carefully designed form of high-risk pooling can significantly reduce the undercompensation of chronically ill patients, while retaining a significant portion of the incentives for efficient health care procurement. A redistribution of 2.6% of the total risk equalization budget results in a 42% reduction in selection incentives. Thus, high-risk pooling serves as a complementary tool to risk equalization and a promising alternative to other forms of risk sharing.

Chapter 6 takes an even more fundamental approach by questioning not the risk equalization system itself, but rather the regulation of health insurance premiums. Within regulated health insurance markets, it is generally assumed that uniform premiums are necessary to ensure solidarity between healthy and unhealthy individuals. At the same time, it is precisely this premium regulation that creates incentives for selection, because health insurers are not allowed to charge premiums that reflect the remaining predictable differences in health care costs or the residual spending after equalization contributions.

This chapter examines the extent to which the advanced Dutch risk equalization system already achieves the desired cross-subsidies between healthy and unhealthy insured individuals. It then simulates the consequences that a partial relaxation of premium regulation would have for both the magnitude of cross-subsidies and the remaining selection incentives. The results show that a substantial portion of solidarity is already achieved through risk equalization. This creates a policy trade-off between the additional solidarity achieved by uniform premiums, on the one hand, and the selection incentives caused precisely by this regulation, on the other. The chapter thus demonstrates that the institutional design of regulated competition itself may be subject to reconsideration when risk equalization has reached a high level of development but still leaves room for selection incentives.

The findings of Chapters 4, 5, and 6 make it clear that no single policy measure exists that can completely eliminate selection incentives. Further improvements to risk equalization, additional forms of risk sharing, and adjustments to the regulation of health insurance markets each appear capable of contributing to the reduction of selection incentives, but

at the same time involve their own trade-offs regarding efficiency, solidarity, feasibility, and financial accessibility. Taken together, the three chapters demonstrate that reducing selection incentives requires a comprehensive approach, in which various policy instruments complement rather than replace one another.

Policy Implications

The studies in this dissertation show that, despite the significant developments that risk equalization models have undergone in recent decades, financial incentives for risk selection persist even within the most advanced systems. The Dutch risk equalization model ranks among the most sophisticated models internationally, but it appears unable to fully compensate for all predictable differences in health care costs among insured individuals. As a result, health insurers continue to face predictable profits and losses for specific groups of insured individuals, which perpetuates the incentives for risk selection. The findings of this dissertation have several implications for policymakers responsible for designing regulated health insurance markets and risk equalization systems.

First, the results emphasize the importance of a broader evaluation of risk equalization. Traditional evaluation studies typically focus on average undercompensation and overcompensation for statically defined groups of insured individuals. The analyses in this dissertation demonstrate that selection incentives are also influenced by the dynamics of health changes, the timing of compensation, the behavior of insured individuals in response to changes in their health status, and the uncertainty associated with financial outcomes after risk equalization. Incorporating these additional dimensions into the evaluation provides a more complete picture of the selection incentives that health insurers face in practice. It is therefore recommended that these aspects be systematically included in evaluations of risk equalization models.

A second policy implication is that further improvements to risk equalization remain possible, even within already advanced models such as the Dutch system. Further refinement of the risk equalization model could lead to more accurate compensation for specific groups of chronically ill patients. At the same time, the results illustrate that further model expansions yield increasingly smaller improvements, while the complexity of the system increases. Policymakers will therefore need to weigh which improvements, given the available data sources, offer sufficient societal added value relative to the additional administrative burdens.

Third, this dissertation shows that reducing selection incentives does not have to occur exclusively through further expansion of risk equalization. A carefully designed system of high-risk pooling can further limit the financial risks for health insurers, while preserving important efficiency incentives. Such instruments offer opportunities to reduce persistent undercompensation of specific groups of chronically ill patients without fundamentally

altering the functioning of the risk equalization system. The final design always requires a careful balance between limiting selection incentives and maintaining incentives for efficient health care procurement.

A fourth policy implication concerns the relationship between risk equalization and premium regulation. The analyses suggest that a substantial portion of the desired solidarity is already achieved through risk equalization. As a result, the traditional balance between uniform premiums and risk equalization warrants renewed attention. As risk equalization explicitly achieves a larger share of the desired solidarity, there is scope to reassess the balance between premium regulation, solidarity, and selection incentives. This does not mean that premium regulation should be abandoned, but rather that the social costs and benefits of premium regulation can be reevaluated in light of the current quality of risk equalization.

At the same time, the results make it clear that none of the policy strategies examined, taken individually, eliminates all selection incentives. Individually, the measures reduce a specific aspect of the problem, but at the same time introduce new trade-offs in terms of efficiency, solidarity, feasibility, and financial accessibility. Reducing selection incentives therefore does not require a single optimal policy measure, but rather a careful combination of instruments that collectively contribute to a balanced structure of regulated health insurance markets.

In this way, this dissertation underscores that risk equalization should not be viewed as a static instrument, but as a system that must be continuously adapted to changes in medical care, available data, and the behavioral responses of health insurers and insured individuals. Only by systematically evaluating risk equalization on a periodic basis and further developing it where necessary can the system of regulated competition continue to sustainably support both solidarity and efficient competition.

Recommendations for Further Research

The results of this dissertation offer several starting points for future research into the further development of risk equalization and regulated health insurance markets.

A first recommendation concerns the further development of evaluation methods for risk equalization. This dissertation shows that selection incentives cannot be fully captured using average undercompensation and overcompensation for statically defined groups of insured individuals. Future research can broaden this evaluation by paying closer attention to the dynamics of health, the behavior of insured individuals and health insurers, the uncertainty surrounding financial outcomes, and the interaction between these factors. This can provide an even more complete picture of how selection incentives manifest themselves in practice.

A second recommendation concerns the further development of risk equalization models. Although this dissertation shows that additional health information from primary care can improve compensation for specific groups of chronically ill patients, research into new data sources and innovative equalization features remains necessary. The increasing availability of electronic health records and other health data offers opportunities to further refine risk equalization models. In doing so, it remains important to focus not only on the predictive power of models, but also on feasibility, administrative burdens, and the efficiency incentives generated by the model.

In addition, research into alternative policy instruments deserves further attention. The analyses in this dissertation show that risk equalization is not the only instrument that can be used to reduce selection incentives. Both innovative forms of risk sharing and adjustments to the ban on premium differentiation can offer additional possibilities. Future research could focus on combinations of such policy instruments and the question of under what circumstances they reinforce or, conversely, counteract one another.

A fourth recommendation concerns the international applicability of the findings. Although the empirical analyses in this dissertation were conducted within the Dutch context, the underlying mechanisms surrounding risk equalization, risk selection, and regulated competition are also relevant to other countries with comparable systems. Comparative international research can contribute to a better understanding of the extent to which the evaluation methods and policy strategies presented here are transferable to other institutional contexts.

Finally, the results of this dissertation underscore that risk equalization is an ongoing process. As risk equalization models continue to be refined, the nature of the remaining selection incentives and the associated policy questions also change. This calls for a continuous process of evaluation, model development, and policy analysis. The central message of this dissertation is therefore that reducing selection incentives requires a continuous interaction between scientific research and policy development. Only through this interaction can the balance between solidarity, accessibility, efficiency, and competition be sustainably maintained within regulated health insurance markets.

SAMENVATTING

SAMENVATTING

Aanleiding van het onderzoek

Een systeem van zorgverzekeringen kan een belangrijke rol vervullen in de financiële toegankelijkheid van gezondheidszorg. Echter, zonder regulering worden premies aangepast naar de verwachte zorgkosten van individuele verzekerden, waardoor met name individuen met een verhoogd gezondheidsrisico geconfronteerd worden met hoge of zelfs onbetaalbare premies. Om de toegankelijkheid en solidariteit binnen het zorgstelsel te waarborgen, hebben verschillende landen, waaronder Nederland, Duitsland, Zwitserland en de Verenigde Staten, stelsels ingericht waarin concurrentie tussen zorgverzekeraars wordt gecombineerd met uitgebreide overheidsregulering. Kenmerkende elementen van deze regulering zijn onder meer een acceptatieplicht, een gestandaardiseerd basispakket en beperkingen ten aanzien van premiedifferentiatie.

Hoewel deze combinatie van regulering de betaalbaarheid en toegankelijkheid van zorgverzekeringen beschermt, creëert zij tegelijkertijd financiële prikkels voor zorgverzekeraars om verzekerden met een laag gezondheidsrisico aan te trekken en verzekerden met een hoog gezondheidsrisico juist te vermijden. Wanneer alle verzekerden immers eenzelfde premie betalen, maar voorspelbaar verschillen in de zorgkosten die zij genereren, ontstaan prikkels voor risicoselectie door zorgverzekeraars. Zij kunnen hier vervolgens op acteren in de vormgeving van verzekeringsproducten en marketingstrategieën of bijvoorbeeld via selectieve contractering van zorgaanbieders. Risicoselectie kan daarmee de toegankelijkheid, kwaliteit en doelmatigheid van de gezondheidszorg onder druk zetten en de eerlijke concurrentie tussen zorgverzekeraars verstoren.

Om deze prikkels te beperken wordt in verschillende stelsels met gereguleerde concurrentie het instrument van risicoverevening ingezet. Het risicovereveningssysteem beoogt zorgverzekeraars vooraf een risicoafhankelijke compensatie te geven voor de voorspelbare verschillen in zorgkosten van hun verzekerden. Naarmate het risicovereveningsmodel beter in staat is om verschillen in verwachte zorgkosten te voorspellen, nemen de voorspelbare winsten en verliezen op specifieke groepen verzekerden af en verminderen daarmee ook de prikkels tot risicoselectie. De afgelopen decennia zijn de toegepaste modellen daarom uitgegroeid van relatief eenvoudige modellen, gebaseerd op demografische kenmerken, tot geavanceerde modellen waarin gebruik wordt gemaakt van diepgaande informatie over gezondheid en zorggebruik.

Ondanks deze voortdurende verbeteringen blijkt uit zowel nationale als internationale onderzoeken dat ook de meest geavanceerde risicovereveningsmodellen nog steeds leiden tot substantiële onder- en overcompensaties voor specifieke groepen verzekerden, met name voor individuen met chronische aandoeningen. Hierdoor blijven financiële prikkels tot risicoselectie bestaan. Verdere verbetering van risicoverevening blijft daarom

noodzakelijk om de toegankelijkheid, solidariteit en doelmatigheid van zorgverzekeringsmarkten te beschermen.

Het reduceren van selectieprikkels vereist enerzijds een zorgvuldige evaluatie van de mate waarin risicoverevening er daadwerkelijk in slaagt voorspelbare kostenverschillen tussen verzekerden te compenseren. Anderzijds is het noodzakelijk om aanvullende strategieën te ontwikkelen en te toetsen die de resterende selectieprikkels verder kunnen verminderen. Dit proefschrift draagt aan beide doelstellingen bij door nieuwe methoden te ontwikkelen voor het evalueren van risicoverevening en door verschillende beleidsopties te onderzoeken die het functioneren van gereguleerde zorgverzekeringsmarkten verder kunnen verbeteren.

Doel van het onderzoek

Het doel van dit proefschrift betreft daarom het ontwikkelen van nieuwe methoden om de selectieprikkels die na risicoverevening blijven bestaan beter te evalueren en het onderzoeken van strategieën om deze selectieprikkels verder te verminderen. Dit proefschrift beoogt bij te dragen aan de verdere ontwikkeling van risicoverevening door zowel nieuwe evaluatiemethoden te introduceren als verschillende beleidsstrategieën te ontwikkelen en empirisch te toetsen.

Het eerste deel van het proefschrift richt zich op het evalueren van de werking van risicoverevening met behulp van innovatieve methoden. Hiervoor worden nieuwe benaderingen ontwikkeld die gebruikmaken van objectieve huisartsgegevens en administratieve gegevens om de effectiviteit van risicoverevening vanuit nieuwe perspectieven te beoordelen. Het eerste deel van het proefschrift heeft daarmee tot doel het inzicht in de aard, omvang en dynamiek van resterende selectieprikkels te vergroten.

Het tweede deel van het proefschrift richt zich op het simuleren van strategieën die deze selectieprikkels verder kunnen verminderen. De eerste strategie betreft het verbeteren van een bestaand kenmerk van het Nederlandse risicovereveningsmodel. De tweede strategie onderzoekt de mogelijkheden van een innovatieve vorm van risicodeling als aanvulling op risicoverevening. De derde strategie verkent in hoeverre een gedeeltelijke versoepeling van premieregulering kan bijdragen aan het verminderen van selectieprikkels. Door deze drie complementaire strategieën te analyseren wordt inzicht verkregen in de afwegingen voor beleidsmakers ten aanzien van het beperken van risicoselectie, het behouden van doelmatigheidsprikkels, het waarborgen van solidariteit en het bevorderen van de toegankelijkheid van de zorg.

Deel I: Het evalueren van risicoverevening door innovatieve methoden

Hoewel risicovereveningsmodellen doorgaans worden beoordeeld aan de hand van de voorspelbare winsten en verliezen voor specifieke groepen verzekerden, kennen de gangbare evaluatiemethoden verschillende beperkingen. Zo wordt veelal uitgegaan van één observatiejaar en worden selectieprikkels uitsluitend beoordeeld op basis van gemiddelde onder- en overcompensaties voor afgebakende groepen verzekerden. Hierdoor blijven belangrijke aspecten van selectieprikkels buiten beeld. Het eerste deel van dit proefschrift introduceert daarom twee nieuwe benaderingen die de evaluatie van risicoverevening verbreden. Daarbij wordt gebruikgemaakt van objectieve huisartsgegevens uit de Nivel Primary Care Database en administratieve gegevens uit het Nederlandse risicovereveningssysteem.

Hoofdstuk 2 onderzoekt in hoeverre prospectieve risicoverevening zorgverzekeraars adequaat compenseert voor verzekerden die worden gediagnosticeerd met een nieuwe chronische aandoening. Eerdere evaluatiestudies beschouwen personen met een chronische aandoening veelal als één homogene groep, terwijl het moment waarop een aandoening ontstaat naar verwachting van grote invloed op de gerelateerde zorgkosten en de mate waarin zorgverzekeraars worden gecompenseerd. Omdat prospectieve risicoverevening gebruikmaakt van informatie uit voorgaande jaren, kan een plotselinge verslechtering van de gezondheid niet direct worden verwerkt in de compensatie die zorgverzekeraars ontvangen.

Met behulp van huisartsgegevens worden personen geïdentificeerd die in een bepaald jaar voor het eerst worden gediagnosticeerd met een chronische aandoening. Vervolgens worden hun zorgkosten en de vereveningsbijdragen gevolgd vanaf twee jaar vóór tot en met twee jaar na de diagnose. Daarnaast wordt onderzocht in hoeverre deze verzekerden rondom het jaar van diagnose van zorgverzekeraar wisselen, aangezien dergelijke overstappen bepalend kunnen zijn voor risicoselectie door zorgverzekeraars. De resultaten laten zien dat zorgverzekeraars in het jaar waarin een chronische aandoening voor het eerst wordt vastgesteld met substantiële verliezen worden geconfronteerd. Ook in de jaren vóór en na de diagnose blijken voor verschillende aandoeningen nog aanzienlijke ondercompensaties te bestaan. Tegelijkertijd blijkt dat verzekerden reageren op de verandering in hun gezondheidsstatus bij het kiezen van een zorgverzekeraar. De combinatie van deze bevindingen suggereert dat de ondercompensaties kunnen leiden tot belangrijke selectieprikkels voor zorgverzekeraars. De studie laat daarmee zien dat het meenemen van de dynamiek rondom het ontstaan van chronische aandoeningen nieuwe inzichten oplevert in selectieprikkels en verdere verbetering van het vereveningsmodel noodzakelijk blijft.

Hoofdstuk 3 introduceert vervolgens een tweede innovatieve benadering voor het evalueren van risicoverevening. Waar bestaande evaluaties zich vrijwel uitsluitend richten op de gemiddelde voorspelbare winsten en verliezen van verzekerden, onderzoekt dit hoofdstuk een andere bron van selectieprikkels: de spreiding van de resterende zorgkosten na bijdragen vanuit de risicoverevening. Zelfs wanneer zorgverzekeraars gemiddeld genomen volledig worden gecompenseerd voor de zorgkosten van verschillende groepen verzekerden, kunnen aanzienlijke verschillen bestaan in de onzekerheid rondom de uiteindelijke financiële resultaten van zorgverzekeraars. In dit hoofdstuk wordt aangetoond dat de residuele zorgkosten na risicoverevening heteroscedastisch zijn: groepen met hogere verwachte zorgkosten kennen doorgaans ook een grotere spreiding in de resterende financiële resultaten voor zorgverzekeraars. Voor risicomijdende zorgverzekeraars betekent deze grotere onzekerheid dat het contracteren van dergelijke verzekerden minder aantrekkelijk kan zijn, ook wanneer de gemiddelde financiële uitkomst neutraal is. Daarnaast leidt het tot hogere kapitaalvereisten vanuit het toezicht op de solvabiliteit van zorgverzekeraars. Door deze mechanismen kan heteroscedasticiteit een zelfstandige bron van selectieprikkels vormen die in de bestaande literatuur grotendeels buiten beschouwing is gebleven.

De bevindingen van hoofdstuk 2 en 3 laten gezamenlijk zien dat selectieprikkels niet volledig worden gevangen door uitsluitend te kijken naar de gemiddelde onder- en overcompensaties voor statisch gedefinieerde groepen verzekerden. Zowel de ontwikkeling van de gezondheidstoestand van verzekerden in de tijd als de onzekerheid rondom de financiële uitkomsten na risicoverevening blijken aanvullende dimensies te vormen die relevant zijn voor het evalueren van de werking van risicoverevening. Daarmee verbreedt het eerste deel van dit proefschrift het instrumentarium waarmee de effectiviteit van risicovereveningssystemen kan worden beoordeeld en ontstaat een completer beeld van de selectieprikkels die ondanks geavanceerde risicoverevening blijven bestaan.

Deel II: Het simuleren van strategieën om selectieprikkels te verminderen

Waar het eerste deel van dit proefschrift zich richt op het beter begrijpen en evalueren van de resterende selectieprikkels na risicoverevening, staat in het tweede deel de vraag centraal hoe deze selectieprikkels verder kunnen worden verminderd. Dit deel van het proefschrift onderzoekt daarom drie complementaire beleidsstrategieën om deze prikkels verder te beperken. De eerste strategie richt zich op het verbeteren van het risicovereveningsmodel zelf, de tweede op het aanvullen van risicoverevening met een innovatieve vorm van risicodeling, en de derde op het heroverwegen van de regulering ten aanzien van premiedifferentiatie.

Hoofdstuk 4 onderzoekt in hoeverre selectieprikkels kunnen worden verminderd door een verdere verbetering van één van de gezondheidsgerelateerde vereveningsken-

merken in het Nederlandse risicovereveningsmodel. Het kenmerk diagnosekostengroepen is ingericht om op basis van ziekenhuisdeclaraties de aanwezigheid van chronische aandoeningen te identificeren en zorgverzekeraars vervolgens gericht te compenseren. Het doel van de aanpassing van het kenmerk is om, via het toepassen van een nieuwe clustertechniek van de onderliggende diagnoses en behandelingen en het toestaan van meervoudige indeling van verzekerden in het kenmerk, de nauwkeurigheid van het vereveningsmodel te verbeteren. Om deze verbetering objectief te evalueren worden huisartsgegevens uit de Nivel Primary Care Database ingezet om de aanwezigheid van een chronische aandoening vast te stellen en vervolgens te toetsen hoe de modelverbetering uitwerkt voor de te onderscheiden groepen.

De resultaten laten zien dat de modelaanpassing leidt tot een aanzienlijke afname van voorspelbare winsten en verliezen voor groepen verzekerden die worden geïdentificeerd met de declaratiegegevens en verzekerden met multimorbiditeit. Echter, voor de groepen verzekerden die worden geïdentificeerd met behulp van de huisartsengegevens zijn de resultaten minder overtuigend. Dit toont aan dat, ondanks de verbetering van het model, verzekerden met een chronische aandoening grotendeels ondergecompenseerd blijven door het vereveningsmodel. Verdere verbetering van het model is daarom noodzakelijk om de efficiëntie en toegankelijkheid voor deze kwetsbare groepen en het zorgsysteem als geheel te versterken.

In **Hoofdstuk 5** is vervolgens een duidelijk andere benadering onderzocht. In plaats van het risicovereveningsmodel verder uit te breiden, wordt een innovatieve vorm van risicodeling – high-risk pooling – gesimuleerd als aanvulling op de risicoverevening. Binnen de systematiek van high-risk pooling worden vooraf geselecteerde verzekerden aan een aparte pool toegewezen als wordt verwacht dat zij uitzonderlijk hoge resterende zorgkosten hebben na het toekennen van de vereveningsbijdragen. De verzekeraars van deze individuen worden vervolgens via een gerichte bijdrage gecompenseerd voor de uiteindelijke mismatch tussen diens zorgkosten en vereveningsbijdrage. Binnen dit onderzoek worden verschillende strategieën op historische data toegepast om een groep hoogrisico verzekerden te identificeren. Vervolgens is een selectie gemaakt van de omvang van de pool om de selectieprikkels gericht te reduceren en doelmatigheidsprikkels zo veel mogelijk in stand te houden.

In dit hoofdstuk worden verschillende varianten van high-risk pooling en alternatieve vormen van risicodeling ontwikkeld en met elkaar vergeleken. De analyses laten zien dat een zorgvuldig ontworpen vorm van high-risk pooling de ondercompensatie van chronisch zieken aanzienlijk kan verminderen, terwijl een belangrijk deel van de prikkels voor doelmatige zorginkoop behouden blijft. Bij een herverdeling van 2.6% van het totale budget voor de risicoverevening wordt een reductie van selectieprikkels van

42% bewerkstelligd. Daarmee vormt high-risk pooling een aanvullend instrument naast risicoverevening en een veelbelovend alternatief voor andere vormen van risicodeling.

Hoofdstuk 6 kiest een nog fundamentelere invalshoek door niet het risicovereveningssysteem zelf, maar de regulering van zorgverzekeringspremies ter discussie te stellen. Binnen gereguleerde zorgverzekeringsmarkten wordt doorgaans verondersteld dat uniforme premies noodzakelijk zijn om solidariteit tussen gezonde en ongezonde verzekerden te waarborgen. Tegelijkertijd vormt juist deze premieregulering de aanleiding voor selectieprikkels, doordat zorgverzekeraars geen premie mogen vragen die aansluit bij de resterende voorspelbare verschillen in zorgkosten of het restant na vereveningsbijdragen.

In dit hoofdstuk wordt onderzocht in hoeverre het geavanceerde Nederlandse risicovereveningssysteem reeds een belangrijk deel van de gewenste kruissubsidies tussen gezonde en ongezonde verzekerden realiseert. Vervolgens wordt gesimuleerd welke gevolgen een gedeeltelijke versoepeling van premieregulering zou hebben voor zowel de omvang van kruissubsidies als de resterende selectieprikkels. De resultaten laten zien dat een substantieel deel van de solidariteit reeds via risicoverevening tot stand komt. Hierdoor ontstaat een beleidsmatige afweging tussen de aanvullende solidariteit die uniforme premies realiseren enerzijds en de selectieprikkels die juist door deze regulering worden veroorzaakt anderzijds. Het hoofdstuk laat daarmee zien dat ook de institutionele vormgeving van gereguleerde concurrentie zelf onderwerp van heroverweging kan zijn wanneer risicoverevening een hoog ontwikkelingsniveau heeft bereikt, maar ruimte over laat voor selectieprikkels.

De bevindingen van hoofdstuk 4, 5 en 6 maken duidelijk dat er geen afzonderlijke beleidsmaatregel bestaat waarmee selectieprikkels volledig kunnen worden weggenomen. Verdere verbetering van risicoverevening, aanvullende vormen van risicodeling en aanpassingen in de regulering van zorgverzekeringsmarkten blijken ieder een bijdrage te kunnen leveren aan het verminderen van selectieprikkels, maar gaan tegelijkertijd gepaard met eigen afwegingen ten aanzien van doelmatigheid, solidariteit, uitvoerbaarheid en financiële toegankelijkheid. Gezamenlijk laten de drie hoofdstukken zien dat het verminderen van selectieprikkels vraagt om een integrale benadering, waarin verschillende beleidsinstrumenten elkaar aanvullen in plaats van vervangen.

Beleidsimplicaties

De onderzoeken in dit proefschrift laten zien dat, ondanks de aanzienlijke ontwikkeling die risicovereveningsmodellen de afgelopen decennia hebben doorgemaakt, financiële prikkels tot risicoselectie ook binnen de meest geavanceerde systemen blijven bestaan. Het Nederlandse risicovereveningsmodel behoort internationaal tot de meest verfijnde modellen, maar blijkt niet in staat alle voorspelbare verschillen in zorgkosten tussen

verzekerden volledig te compenseren. Hierdoor blijven zorgverzekeraars voor bepaalde groepen verzekerden geconfronteerd worden met voorspelbare winsten en verliezen, hetgeen de prikkels tot risicoselectie in stand houdt. De bevindingen van dit proefschrift hebben verschillende implicaties voor beleidsmakers die verantwoordelijk zijn voor de inrichting van gereguleerde zorgverzekeringsmarkten en risicovereveningssystemen.

In de eerste plaats benadrukken de resultaten het belang van een bredere evaluatie van risicoverevening. Traditionele evaluatiestudies richten zich veelal op gemiddelde onderen overcompensaties voor statisch gedefinieerde groepen verzekerden. De analyses in dit proefschrift tonen aan dat selectieprikkels daarnaast worden beïnvloed door de dynamiek waarmee gezondheid zich ontwikkelt, de timing van compensatie, het gedrag van verzekerden rondom veranderingen in hun gezondheidstoestand en de onzekerheid waarmee financiële resultaten na risicoverevening gepaard gaan. Door deze aanvullende dimensies in de evaluatie te betrekken ontstaat een vollediger beeld van de selectieprikkels waarmee zorgverzekeraars in de praktijk worden geconfronteerd. Het verdient daarom aanbeveling deze aspecten structureel te betrekken bij evaluaties van risicovereveningsmodellen.

Een tweede beleidsimplicatie is dat verdere verbeteringen van risicoverevening mogelijk blijven, ook binnen reeds geavanceerde modellen zoals het Nederlandse systeem. Verdere verfijning van het risicovereveningsmodel kan leiden tot een nauwkeurigere compensatie van specifieke groepen chronisch zieken. Tegelijkertijd illustreren de resultaten dat verdere modeluitbreidingen steeds kleinere verbeteringen opleveren, terwijl de complexiteit van het systeem toeneemt. Beleidsmakers zullen daarom moeten afwegen welke verbeteringen, gegeven de beschikbare databronnen, voldoende maatschappelijke meerwaarde bieden ten opzichte van de extra uitvoeringslasten.

Ten derde laat dit proefschrift zien dat het verminderen van selectieprikkels niet uitsluitend hoeft plaats te vinden via verdere uitbreiding van risicoverevening. Een zorgvuldig vormgegeven systeem van high-risk pooling kan de financiële risico's voor zorgverzekeraars verder beperken, terwijl belangrijke doelmatigheidsprikkels behouden blijven. Dergelijke instrumenten bieden mogelijkheden om hardnekkige ondercompensaties van specifieke groepen chronisch zieken te verminderen zonder de werking van het risicovereveningssysteem fundamenteel te wijzigen. De uiteindelijke vormgeving vraagt daarbij steeds om een zorgvuldige afweging tussen het beperken van selectieprikkels en het behouden van prikkels voor doelmatige zorginkoop.

Een vierde beleidsimplicatie betreft de samenhang tussen risicoverevening en premieregulering. De analyses suggereren dat een substantieel deel van de gewenste solidariteit reeds via risicoverevening wordt gerealiseerd. Hierdoor verdient de traditionele balans tussen uniforme premies en risicoverevening hernieuwde aandacht. Naarmate

risicoverevening een groter deel van de gewenste solidariteit expliciet realiseert, ontstaat ruimte om de balans tussen premieregulering, solidariteit en selectieprikkels opnieuw te beoordelen. Dit betekent niet dat premieregulering zou moeten worden losgelaten, maar wel dat de maatschappelijke kosten en baten van premieregulering opnieuw kunnen worden afgewogen in het licht van de huidige kwaliteit van risicoverevening.

Tegelijkertijd maken de resultaten duidelijk dat geen van de onderzochte beleidsstrategieën afzonderlijk alle selectieprikkels wegneemt. De maatregelen verminderen individueel een specifiek onderdeel van het probleem, maar introduceren tegelijkertijd nieuwe afwegingen op het gebied van doelmatigheid, solidariteit, uitvoerbaarheid en financiële toegankelijkheid. Het terugdringen van selectieprikkels vraagt daarom niet om één optimale beleidsmaatregel, maar om een zorgvuldige combinatie van instrumenten die gezamenlijk bijdragen aan een evenwichtige inrichting van gereguleerde zorgverzekeringsmarkten.

Daarmee onderstreept dit proefschrift dat risicoverevening niet moet worden beschouwd als een statisch instrument, maar als een systeem dat voortdurend moet worden aangepast aan veranderingen in medische zorg, beschikbare gegevens en de gedragsreacties van zorgverzekeraars en verzekerden. Alleen door risicoverevening structureel periodiek te evalueren en waar nodig verder te ontwikkelen kan het stelsel van gereguleerde concurrentie zowel solidariteit als doelmatige concurrentie duurzaam blijven ondersteunen.

Aanbevelingen voor vervolgonderzoek

De resultaten van dit proefschrift bieden verschillende aanknopingspunten voor toekomstig onderzoek naar de verdere ontwikkeling van risicoverevening en gereguleerde zorgverzekeringsmarkten.

Een eerste aanbeveling betreft de verdere ontwikkeling van evaluatiemethoden voor risicoverevening. Dit proefschrift laat zien dat selectieprikkels niet volledig kunnen worden gevangen aan de hand van gemiddelde onder- en overcompensaties voor statisch gedefinieerde groepen verzekerden. Toekomstig onderzoek kan deze evaluatie verder verbreden door meer aandacht te besteden aan de dynamiek van gezondheid, het gedrag van verzekerden en zorgverzekeraars, de onzekerheid rondom financiële uitkomsten en de wisselwerking tussen deze factoren. Daarmee kan een nog completer beeld ontstaan van de wijze waarop selectieprikkels zich in de praktijk manifesteren.

Een tweede aanbeveling betreft het verder ontwikkelen van risicovereveningsmodellen. Hoewel dit proefschrift laat zien dat aanvullende gezondheidsinformatie uit de eerstelijnszorg de compensatie van specifieke groepen chronisch zieken kan verbeteren, blijft onderzoek naar nieuwe databronnen en innovatieve vereveningskenmerken noodzakelijk. De toenemende beschikbaarheid van elektronische patiëntendossiers en andere

gezondheidsgegevens biedt mogelijkheden om risicovereveningsmodellen verder te verfijnen. Daarbij blijft het van belang om niet uitsluitend aandacht te besteden aan de voorspelkracht van modellen, maar ook aan uitvoerbaarheid, administratieve lasten en de doelmatigheidsprikkels die van het model uitgaan.

Daarnaast verdient onderzoek naar alternatieve beleidsinstrumenten verdere aandacht. De analyses in dit proefschrift laten zien dat risicoverevening niet het enige instrument is waarmee selectieprikkels kunnen worden verminderd. Zowel innovatieve vormen van risicodeling als aanpassingen aan het verbod op premiedifferentiatie kunnen aanvullende mogelijkheden bieden. Toekomstig onderzoek kan zich richten op combinaties van dergelijke beleidsinstrumenten en de vraag onder welke omstandigheden zij elkaar versterken of juist tegenwerken.

Een vierde aanbeveling betreft de internationale toepasbaarheid van de bevindingen. Hoewel de empirische analyses in dit proefschrift zijn uitgevoerd binnen de Nederlandse context, zijn de onderliggende mechanismen rondom risicoverevening, risicoselectie en gereguleerde concurrentie ook relevant voor andere landen waarin vergelijkbare stelsels bestaan. Vergelijkend internationaal onderzoek kan bijdragen aan een beter begrip van de mate waarin de gepresenteerde evaluatiemethoden en beleidsstrategieën overdraagbaar zijn naar andere institutionele contexten.

Tot slot onderstrepen de resultaten van dit proefschrift dat risicoverevening geen eindpunt kent. Naarmate risicovereveningsmodellen verder worden verbeterd, veranderen ook de aard van de resterende selectieprikkels en de beleidsvragen die daarmee samenhangen. Dit vraagt om een continu proces van evaluatie, modelontwikkeling en beleidsanalyse. De centrale boodschap van dit proefschrift is daarom dat het verminderen van selectieprikkels vraagt om een voortdurende wisselwerking tussen wetenschappelijk onderzoek en beleidsontwikkeling. Alleen door deze wisselwerking kan het evenwicht tussen solidariteit, toegankelijkheid, doelmatigheid en concurrentie duurzaam worden behouden binnen gereguleerde zorgverzekeringsmarkten.

DANKWOORD

DANKWOORD

Kenmerkend voor mijn handelen schrijf ik dit dankwoord onder de nodige tijdsdruk. Enig uitstelgedrag heeft ervoor gezorgd dat ik voor het einde van de week mijn dankwoord moet leveren aan de drukker, terwijl de eerste woorden nu pas op het digitale papier verschijnen. Ach, zoals ik mijzelf doorgaans, ietwat naïef en bagatelliserend, voorhoud: *diamonds are made under pressure*. Uiteindelijk komt alles wel goed en ook dát is tekenend voor het proces van dit gehele promotietraject. Typerend was ook mijn gevoel bij de (achteraf terechte) waarschuwing van Richard en Erik toen ik aankondigde een nieuwe uitdaging buiten de academische sector aan te gaan en het proefschrift wel even in de avond- en weekenduren af te ronden. Dat afronden zou uiteindelijk nog bijna twee jaar duren, maar het is gelukt. Dat deze woorden nu rechtstreeks uit het fysieke bewijs daarvan gelezen kunnen worden, is mede te danken aan verschillende mensen. Graag richt ik hier een woord aan hen die belangrijk voor mij zijn geweest tijdens dit buitengewoon leerzame traject.

Allereerst richt ik mij graag tot de twee personen die, als het gaat om de inhoud van dit proefschrift, centraal staan: Richard van Kleef en Erik Schut. Het schrijven van dit werk was onmogelijk zonder de waardevolle begeleiding van jullie beiden.

Erik, graag wil ik je bedanken voor het nemen van de tijd en energie om mijn stukken te beoordelen. Hoewel de dagelijkse begeleiding meer voor rekening van Richard kwam, is jouw kritische blik in de latere stadia van de onderzoeken en het afronden van het proefschrift als geheel van belangrijke meerwaarde geweest. Daarnaast wil ik je graag bedanken voor de kans om naast het uitvoeren van wetenschappelijk onderzoek naar risicoverevening ook in de praktijk vlieguren op het onderwerp te maken. Deze ervaring heeft ervoor gezorgd dat ik ook in mijn carrière buiten de universiteit, doch binnen het domein van de risicoverevening, naadloos kon instromen.

Richard, hoe breng ik ooit onder woorden wat jij voor mij hebt betekend. In de zomer van 2019 kwam ik naar de Erasmus Universiteit voor een luchtig gesprek over jouw idee voor een promotieonderzoek. Althans, dat was mijn inschatting. Het gesprek verliep zodanig dat je na afloop mij de kans bood om het onderzoek ook daadwerkelijk uit te voeren. Niet eerder had ik aan promoveren gedacht, maar het enthousiasme waarmee jij over de risicoverevening spreekt zou zelfs een timmerman nog tot het onderwerp overtuigen. Over jouw kennis en vaardigheden op het gebied van zorgverzekeringen en de risicoverevening hoef ik weinig te zeggen. Binnen de verschillende gremia met betrekking tot het onderwerp word je beschouwd als de toonaangevende expert. De professor, zonder professor te zijn. Ik had geen betere leermeester kunnen treffen. Hartelijk dank voor alle feedback, programmatuur en motivatie. Dan is er nog het menselijke aspect. De ontspannen sfeer tijdens dat eerste gesprek tussen ons is typerend geweest voor onze

samenwerking. Jouw geduld, toegankelijkheid en eindeloze vrolijkheid maken je niet alleen een uitstekende begeleider, maar een fantastisch mens. Ik heb andere promovendi gesproken over de begeleiding die zij ontvangen en van de vergelijking die daaruit volgt trek ik één simpele conclusie: er zijn er weinig zoals jij. Dank je wel.

Graag wil ik vervolgens enkele woorden wijden aan de verschillende coauteurs op de onderzoeken die onderdeel uitmaken van dit proefschrift: René van Vliet, Wynand van de Ven en Rudy Douven. Ik ben jullie allen veel dank verschuldigd voor jullie uiterst waardevolle bijdragen aan de verschillende onderzoeken. Jullie uitgebreide kennis over de risicoverevening, zorgverzekeringen en zorgeconomie heeft aanzienlijk bijgedragen aan het schrijven van dit proefschrift. Hartelijk dank voor al jullie commentaar en suggesties op de stukken. Ik heb de samenwerking met jullie altijd als zeer prettig ervaren.

Voorts zijn er nog drie oud-collega's van de vakgroep Health Systems & Insurance (HSI) die ik specifiek wil bedanken. Tadjo, Andreea en Anja. Tadjo en Andreea, wat hebben wij ontzettend gelachen op de universiteit en onze waanzinnige congresreizen. Onderaan de streep hebben we daar weliswaar vakinhoudelijk bijgedragen en mooie onderzoeken bestudeerd, maar zijn we vooral op ontdekking geweest in onder andere Berlijn, Oslo en Kaapstad. Uiteindelijk is dat simpelweg een cruciaal onderdeel van promoveren.

Anja, ook jouw bijdrage wil ik graag uitlichten. Tijdens het schrijven van dit proefschrift heb ik veel aan je gehad. Enerzijds vanwege je uitstekende kennis van het onderwerp en het feit dat ik veel gebruik heb kunnen maken van jouw voorwerk en onderzoeken. Anderzijds omdat je met jouw oprechte interesse in een ander, betrokkenheid en vreugde een prachtig mens bent. Ik ben dankbaar dat ik je bij het ministerie van Volksgezondheid, Welzijn & Sport opnieuw mijn collega mag noemen.

Ook wil ik de andere, niet bij naam genoemde, HSI-collega's bedanken voor de fijne samenwerking van de afgelopen jaren. Velen van jullie hebben meegelezen op mijn stukken en presentaties over de onderzoeken bijgewoond. Hartelijk dank voor jullie feedback hierop.

Verder wil ik het ministerie van Volksgezondheid, Welzijn & Sport, Zorgverzekeraars Nederland en het Nivel bedanken voor het beschikbaar stellen van de data. De empirische analyses binnen dit proefschrift hadden niet uitgevoerd kunnen worden zonder het gebruik van deze administratieve gegevensbestanden. Graag wil ik daarom ook de leden van de begeleidingscommissie wetenschappelijk onderzoek op WOR-data bedanken voor het beschikbaar stellen van deze data en het geven van feedback op de onderzoeken.

Daarnaast wil ik de leden van het Risk Adjustment Network bedanken voor hun uitstekende feedback. Het is een eer om deel uit te hebben gemaakt van dit genootschap van uitmuntende wetenschappers op dit waanzinnige onderwerp. Daarnaast wil ik de redacteuren en reviewers van de tijdschriften bedanken waar de hoofdstukken van dit proefschrift zijn ingediend. Ook ben ik veel dank verschuldigd aan dr. F. Eijkenaar, prof. dr. J. de Jong en prof. dr. J. Boone (beoordelingscommissie) en aan prof. dr. D.M.J. Delnoij, prof. dr. M.C. Mikkers en prof. dr. W.P.M.M. van de Ven (promotiecommissie) voor het lezen en beoordelen van mijn proefschrift en het opponeren bij de verdediging.

Jesse, Jonatham, Nick en Tom, bedankt dat jullie mij soms dwingen om tussen het werk aan het proefschrift door ook dol te zijn. Die belangrijke uitlaatklep helpt om stoom af te blazen en vooral niet al te serieus te worden. Plezier maken staat centraal in het leven. Serieus doen kan altijd nog, later als je groot bent.

Pim, ook wil ik mij graag specifiek tot jou richten. Je bent een fantastische vriend en oprecht geïnteresseerd in het onderwerp van mijn proefschrift, puur en alleen omdat ik mij ermee bezighoud. Zodra ik vol trots mijn gepubliceerde artikelen naar je stuur zeg je het te zullen lezen. Ik denk altijd dat je mij in de maling neemt, maar wanneer we elkaar op een later moment zien haal je daadwerkelijk specifieke aspecten uit het onderzoek aan. Wat een waanzin! Ik waardeer je enorm.

John en Alice, inmiddels zijn jullie een extra setje ouders voor mij geworden. Dank jullie wel voor alles wat jullie voor mij doen, jullie oprechte interesse in mijn werk en jullie kritische vragen over de zaken die ik onderzoek en de bredere aspecten van het leven. John, jij en ik zijn een ware tandem aan het worden. Tijdens onze verbouwing leer jij mij zo ontzettend veel over klussen en in ruil daarvoor ontvang jij kennis over het functioneren van het zorgstelsel. Een eerlijke ruil als je het mij vraagt. Zo wordt deze risicoverevenaar misschien ooit nog eens timmerman en raakt timmerman John tóch ooit in de ban van de risicoverevening.

Ricardo, bedankt dat je bent wie je bent, mijn broer en beste vriend. Jouw oneindige optimisme doet elke tegenslag de das om. Je motiverende woorden zijn geregeld een rots in de branding geweest tijdens dit proces. De strijd tegen risicoselectie gaat onverminderd door, ook na het verdedigen van dit proefschrift.

Lieve pap en mam, jullie hebben ons van jongs af aan de vrijheid gegeven om te doen wat we leuk vinden, als we maar ons best doen. Dankzij jullie kon ik de stappen doorlopen die de start van dit promotietraject voor gingen. Ik wil jullie daarom bedanken voor de fantastische basis die jullie mij gegeven hebben. Ik koester de hechte en warme band die wij hebben. Jullie steun, trots en onvoorwaardelijke liefde hebben een heel belangri-

jke bijdrage geleverd aan het afronden van dit proefschrift. Jullie zijn een voorbeeld voor mij en ik houd enorm veel van jullie.

En dan jij, lieve Roos. De liefde van mijn leven. Ik wil je bedanken voor wie je bent, jouw eindeloze liefde en onvoorwaardelijke steun. Jij maakt mijn wereld zo veel mooier en laat mij genieten van zaken die er echt toe doen. Jij helpt mij de stilte te vinden in elke storm. Ik ben je dankbaar voor hoe jij mij een beter mens maakt, zonder mij te willen veranderen. We werken samen aan het verwezenlijken van onze dromen. Steeds een stukje dichterbij, maar ook vooral genieten van de reis. We gaan nog zo veel mooie dingen meemaken. Lieve Roos, je bent mijn wereld.

Alles komt altijd goed en als het nu nog niet goed is, dan is het nog niet alles.

PhD PORTFOLIO

PHD PORTFOLIO MICHEL OSKAM

Education

Bachelor of Science in 'Beleid en Management van de Gezondheidszorg, Gezondheidswetenschappen', Erasmus University Rotterdam, The Netherlands — 2017

Master of Science in 'European Master in Health Economics & Management' (Joint Degree), Erasmus University Rotterdam, The Netherlands; Management Centre Innsbruck, Austria; University of Bologna, Italy; University of Oslo, Norway — 2019

International peer-reviewed publications

Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2023). Improving Diagnosis-Based Cost Groups in the Dutch Risk Equalization Model: The Effects of a New Clustering Method and Allowing for Multimorbidity. *International Journal of Health Economics and Management*, *23*(2), 303-324. https://doi.org/10.1007/s10754-023-09345-0 — 2023

Oskam, M., Van Kleef, R. C., & Douven, R. (2024). Heteroscedasticity of Residual Spending After Risk Equalization: a Potential Source of Selection Incentives in Health Insurance Markets with premium regulation. *The European Journal of Health Economics*, *25*(3), 379-396. https://doi.org/10.1007/s10198-023-01592-9 — 2024

van Kleef, R. C., van Vliet, R. C., & **Oskam, M.** (2023). Risk Adjustment in Health Insurance Markets: Do Not Overlook the "Real" Healthy. *Medical Care*, *62*(11), 767. https://doi.org/10.1097/MLR.0000000000001955 — 2024

Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2025). Supplementing Risk Adjustment With High-Risk Pooling Using Historical Data for Identifying the High Risks. *Journal of Risk and Insurance*, *92*(1), 166-202. https://doi.org/10.1111/jori.12500 — 2025

Oskam, M., Van Kleef, R. C., & Van Vliet, R. C. (2026). Does Health-Based Prospective Risk Adjustment Adequately Compensate for Individuals Diagnosed With a New Chronic Disease?. *Medical Care Research and Review*, *83*(2), 128-143. https://doi.org/10.1177/10775587251378167 — 2026

Oskam, M., Van Kleef, R. C., & Van de Ven, W. P. M. M. (*Submitted*). To what extent does state-of-the-art risk equalization create cross-subsidies on a competitive health insurance market?. […]

Publications in Dutch journals

Van Kleef, R. C., Van Vliet, R. C. J. A., & **Oskam, M.** (2021). Risicovereveneing tussen zorgverzekeraars nog niet goed genoeg. *De Actuaris*. 2021

Van Kleef, R. C., Van Vliet, R. C. J. A., **Oskam, M.**, & Paturu, A. (2023). Potentiële gamechanger zorgverzekeringsmarkt. *De Actuaris*. 2023

Contributions to Dutch policy research projects

Cattel, D., Eijkenaar, F., **Oskam, M.**, Van Kleef, R. C., Van Vliet, R. C. J. A., & Withagen-Koster, A. A. (2020). Onderzoek risicoverevening 2021: Pre-Overall Toets. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam. 2020

Cattel, D., Eijkenaar, F., **Oskam, M.**, Van Kleef, R. C., Van Vliet, R. C. J. A., & Withagen-Koster, A. A. (2020). Onderzoek risicoverevening 2021: Gegevensfase. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam. 2020

Cattel, D., Eijkenaar, F., **Oskam, M.**, Van Kleef, R. C., Van Vliet, R. C. J. A., & Withagen-Koster, A. A. (2020). Onderzoek risicoverevening 2021: Overall Toets. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam. 2020

Cattel, D., Eijkenaar, F., **Oskam, M.**, Van Kleef, R. C., Van Vliet, R. C. J. A., & Withagen-Koster, A. A. (2020). Onderzoek risicoverevening 2021: Berekening normbedragen. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam. 2020

Van Kleef, R. C., **Oskam, M.**, Visser, L. A., Van Vliet, R. C. J. A., & Withagen-Koster, A. A. (2020). Grootonderhoud DKG's t.b.v. somatisch model 2021 (WOR 988). Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam. 2020

Cattel, D., Eijkenaar, F., **Oskam, M.**, Van Kleef, R. C., & Van Vliet, R. C. J. A. (2021). Onderzoek risicoverevening 2022: Pre-Overall Toets. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam. 2021

Cattel, D., Eijkenaar, F., **Oskam, M.**, Van Kleef, R. C., & Van Vliet, R. C. J. A. (2021). Onderzoek risicoverevening 2022: Gegevensfase. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2021
Cattel, D., Eijkenaar, F., **Oskam, M.**, Van Kleef, R. C., & Van Vliet, R. C. J. A. (2021). Onderzoek risicoverevening 2022: Overall Toets. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2021
Cattel, D., Eijkenaar, F., **Oskam, M.**, Van Kleef, R. C., & Van Vliet, R. C. J. A. (2021). Onderzoek risicoverevening 2022: Berekening normbedragen. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2021
Van Kleef, R. C., Van Vliet, R. C. J. A., & **Oskam, M.** (2021). Restrisico fase III: gezonde verzekerden (WOR 1022). Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2021
Cattel, D., Eijkenaar, F., **Oskam, M.**, Panturu, A., Van Kleef, R. C., & Van Vliet, R. C. J. A. (2022). Onderzoek risicoverevening 2023: Pre-Overall Toets. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2022
Cattel, D., Eijkenaar, F., **Oskam, M.**, Panturu, A., Van Kleef, R. C., & Van Vliet, R. C. J. A. (2022). Onderzoek risicoverevening 2023: Gegevensfase. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2022
Cattel, D., Eijkenaar, F., **Oskam, M.**, Panturu, A., Van Kleef, R. C., & Van Vliet, R. C. J. A. (2022). Onderzoek risicoverevening 2023: Overall Toets. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2022
Cattel, D., Eijkenaar, F., **Oskam, M.**, Panturu, A., Van Kleef, R. C., & Van Vliet, R. C. J. A. (2022). Onderzoek risicoverevening 2023: Berekening normbedragen. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2022
Van Vliet, R. C. J. A., Van Kleef, R. C., & **Oskam, M.** (2022). Berekening normbedragen 2022 op data 2018 (WOR 1054). Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2022
Van Vliet, R. C. J. A., Van Kleef, R. C., & **Oskam, M.** (2022). Verkenning van alternatieven voor herweging van OT-bestanden naar verzekerdenraming. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2022
Cattel, D., Eijkenaar, F., **Oskam, M.**, Panturu, A., Van Kleef, R. C., & Van Vliet, R. C. J. A. (2023). Onderzoek risicoverevening 2024: Pre-Overall Toets. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2023

Cattel, D., Eijkenaar, F., **Oskam, M.**, Panturu, A., Van Kleef, R. C., & Van Vliet, R. C. J. A. (2023). Onderzoek risicoverevening 2024: Gegevensfase. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2023
Cattel, D., Eijkenaar, F., **Oskam, M.**, Panturu, A., Van Kleef, R. C., & Van Vliet, R. C. J. A. (2023). Onderzoek risicoverevening 2024: Overall Toets. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2023
Cattel, D., Eijkenaar, F., **Oskam, M.**, Panturu, A., Van Kleef, R. C., & Van Vliet, R. C. J. A. (2023). Onderzoek risicoverevening 2024: Berekening normbedragen. Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2023
Van Kleef, R. C., Van Vliet, R. C. J. A., **Oskam, M.**, & Panturu, A. (2023). Constrained regression als schattingsmethode voor de risicovereve-ning: mogelijkheden, effecten en afwegingen (WOR 1158). Rapport ESHPM, Rotterdam: Erasmus Universiteit Rotterdam.	2023

PhD training

Searching, finding and managing your literature. Erasmus Universiteit Rotterdam	2019
SAS Base Programming: Fast Track. SAS Institute Huizen	2019
How to finish your PhD in time. Erasmus Universiteit Rotterdam	2019
English Academic Writing for PhD Candidates. Erasmus Universiteit Rotterdam	2019
Professionalism and Integrity in Research. Erasmus Universiteit Rotterdam	2020
Basic Didactics and Group Dynamics. Erasmus Universiteit Rotterdam	2021
Coaching and Intervision for PhD Candidates. Erasmus Universiteit Rotterdam	2022
Data analysis with R. Erasmus Universiteit Rotterdam	2023
Self-presentation: Focus, structure, interaction and visualization. Erasmus Universiteit Rotterdam	2023

Teaching

Statistics-B, pre-master level: teacher	2020 – 2023
Quantitative Research Report, pre-master level: supervisor of students	2020 – 2021
Quantitative Research Report, pre-master level: course coordinator	2021 – 2023
Bachelor thesis supervisor	2020 – 2023

Presentations at conferences and seminars

Commission for the use of Dutch risk equalization data for academic research, The Hague, The Netherlands

- Proposal and outline for the PhD research 2019
- *Improving diagnosis-based cost groups in the Dutch risk equalization model: the effects of a new clustering method and allowing for multimorbidity.* 2020
- *Heteroscedasticity of residual spending after risk equalization: a potential source of selection incentives in health insurance markets with premium regulation.* 2021
- *Supplementing risk adjustment with high-risk pooling using historical data for identifying the high risks.* 2022
- *Does Health-Based Prospective Risk Adjustment Adequately Compensate for Individuals Diagnosed With a New Chronic Disease?* 2023
- *To what extent does state-of-the-art risk equalization create cross-subsidies on a competitive health insurance market?* 2025

Risk Adjustment Network

- Online 2020
- Weggis, Switzerland 2021
- Berlin, Germany 2022
- The Hague, The Netherlands 2023
- Dublin, Ireland 2024

University of Newcastle, Newcastle, Australia (online) 2020

International Health Economics Association

- Online 2020

- Cape Town, South Africa	2023
European Health Economics Association	
- Oslo, Norway	2022
Post Academic Course on Risk Equalization, Rotterdam, The Netherlands	2023
Scientific Platform on Risk Equalization, Rotterdam, The Netherlands	2025

Other

Member of the working group on the research and development of the Dutch risk equalization model	2019 – […]
Organization of the post academic course on risk equalization	2020 – 2023
Member of the team that calculates the risk equalization payment weights for the Dutch Ministry of Health	2020 – 2023
Organization of the ESHPM faculty lustrum event	2022
Organization of the Risk Adjustment Network meeting in The Hague	2023