

Thesis

Alternative estimation methods for high healthcare costs in the Dutch somatic risk equalisation model: a two-part model

Anne van Son
728272

Rotterdam, 5 June 2026
Thesis supervisor: Richard van Kleef
Supervisor Ministry of Health, Welfare and Sport: Lisette Mijnen
Word count: 16,682 words

Acknowledgement

I want to thank my supervisor, Richard van Kleef, for the guidance and constructive feedback during the writing process of this thesis. Furthermore, I would like to thank the Dutch Ministry of Health, Welfare and Sport for providing me with the opportunity to do an internship while writing this thesis. In particular, I would like to thank Lisette Mijnen for her supervision during this internship and for the valuable lessons I learned during my time at the ministry. I am grateful to my family and friends for their encouragement during the writing of this thesis. Finally, I am grateful to the Dutch Ministry of Health, Welfare and Sport and the Association of Health Insurers for access to (anonymised) claims data.

AI-statement

During the writing process of this thesis, AI tools, such as EduGenAI, were used to improve the grammar and clarity of the thesis. The prompts consisted of requesting feedback on individual sentences with unclear wording. No sensitive data was entered into AI systems. All ideas and analyses presented in the thesis are the author's own work.

Summary

The Dutch risk equalisation model undercompensates individuals with high healthcare costs. This undercompensation incentivises insurers to engage in risk selection. In this study, alternative estimation methods were explored to improve the estimation of high healthcare costs. Real-world data on 17 million Dutch individuals with mandatory basic health insurance in 2023 was used for this study.

The aim of this study was to explore two-part models to estimate costs and to assess whether these models perform better in estimating high healthcare costs for somatic care compared to the least squares (LS) regression method. Three two-part models were explored during this study. The models were estimated by combining three cost components. The first component included healthcare costs below the high-cost threshold, while the second component included healthcare costs above the threshold. The third component included the probability of incurring healthcare costs above the threshold. Each two-part model used a different high-cost threshold. The first model defined high costs as costs above €12,400, based on the healthcare costs of the top 5% spenders. The second model defined high costs as costs above €59,874, based on the Pareto tail method. The final two-part model defined high costs as costs above €100,000, based on an absolute monetary threshold obtained from the literature. These models were compared to an LS regression model, which broadly represents the 2026 Dutch risk equalisation model.

The evaluation of the models showed that the LS regression model and the absolute threshold model estimate costs most accurately. The LS regression model performs best for non-high-cost individuals, whereas the absolute threshold model performs best for high-cost individuals. However, the differences in payment fit between these models are relatively small. All models are comparable in terms of efficiency, validity and measurability. However, the two-part models are more complex in comparison to the LS regression model. It remains unclear whether the results from the two-part models are as stable as the results from the LS regression model, as the analyses were not repeated on data from different equalisation years.

This study concludes that the performance of a two-part model is dependent on the threshold which determines high healthcare costs. When applying a fitting threshold, the two-part model improves the payment fit for individuals with high healthcare costs for somatic care relative to the LS model. However, a two-part model, in terms of the payment fit, performs worse for other subgroups. Furthermore, the application of a two-part model is more complex than the use of an LS regression model. This results in a trade-off in which the complexity and payment fit should be evaluated by relevant stakeholders.

Table of contents

1. Problem analysis and research questions	4
2. Theoretical framework	6
2.1 Regulated competition in the Healthcare Insurance Act	6
2.2 The role and relevance of risk equalisation	7
2.3 Methods for estimating healthcare spending	10
2.4 High-cost users	12
2.5 Evaluation measures and the WOR-assessment framework	14
3. Methodology	17
3.1 Data	17
3.2 Data analysis	17
3.3 Evaluation and additional analyses	20
3.4 Ethics, validity and reliability	21
4. Results	22
4.1 Descriptive statistics	22
4.2 The LS regression model	22
4.3 The two-part models	23
4.4 Evaluation and additional analyses	26
5. Discussion	33
5.1 Summary of key findings	33
5.2 Strengths, limitations & directions for future research	35
5.3 Policy implications	37
5.4 Conclusion	38
6. References	39
7. Appendices	43
Appendix 1: Risk adjusters in the 2026 risk equalisation model	43
Appendix 2: Standard evaluation subgroups	44
Appendix 3: Overview of all additional variables in the dataset	45
Appendix 4: Stepwise regression; risk classes	46

1. Problem analysis and research questions

The Dutch healthcare system is based on regulated competition. In a system of regulated competition, insurers supply insurance contracts to their enrollees (Douven, 2000). Because of regulations, insurers are not allowed to ask risk-rated premiums (Douven, 2004). This prohibition creates incentives for risk selection, which provides insurers with incentives to select against individuals with high expected healthcare spending (Oskam et al., 2024). Risk selection creates a barrier to accessing high-value care for individuals with high expected costs (Leao et al., 2023). To counteract the incentives for risk selection, there is a risk equalisation system in the Netherlands. The goal of risk equalisation is to compensate insurers for variation in expected spending among individuals (Van Kleef et al., 2013). Findings by van Kleef and van Vliet (2025) show that certain subgroups are under- or overcompensated by the current risk equalisation model. Furthermore, a study by PwC (2024) concluded that the risk equalisation model strongly undercompensates individuals with high healthcare costs. For this reason, further improvements to the risk equalisation model are necessary to prevent risk selection, particularly for individuals with high healthcare costs. The working group on the development of risk equalisation (WOR) continuously works on improving the Dutch risk equalisation system through the WOR research/policy cycle. The 2025-2026 research program of the WOR does not include research on alternative estimation methods for high healthcare costs (Werkgroep Ontwikkeling Risicoverevening, n.d.). In the absence of improvements for high-cost estimation, the 2026 Dutch risk equalisation model uses ex-post compensation to counteract selection incentives for individuals with high healthcare costs (Ministerie van Volksgezondheid, Welzijn en Sport, 2025a). Such a compensation reimburses insurers after actual costs are realised. This compensation is not ideal, as it reduces an insurer's exposure to risk. By improving the estimation of high healthcare costs, this ex-post compensation might not be necessary anymore.

Research suggests that individuals with high healthcare costs have specific characteristics which differ from those of other individuals (Jansen et al., 2025; Wammes et al., 2017, 2018). This makes it interesting to estimate healthcare costs differently for these individuals. In this study, this will be analysed within the context of the risk equalisation model for somatic care. The undercompensation for individuals with high healthcare costs might be related to the characteristics of the distribution of healthcare spending. A small share of the population has high healthcare costs, while the majority of the population incurs low healthcare costs (Karunayawong et al., 2024). According to Malehi et al. (2015), this skewed distribution is a challenge in modelling healthcare costs. The current equalisation model uses least squares (LS) regression, which is sensitive to skewed distributions and estimates high healthcare costs less accurately. A paper by Kapitulka (2015) states that a two-part model gives a flexible and useful framework for modelling healthcare costs. In this thesis, two-part models will be explored to estimate (high) healthcare costs. This method estimates a conditional outcome using a regression model (Belotti et al., 2015). Two-part models using a zero-cost threshold have been applied previously to this research to predict healthcare costs, for example by Srivastava et al. (2021). However, this study builds on this methodology by implementing a higher cost threshold to predict high healthcare costs. This research, therefore, supplements the currently existing literature on healthcare cost prediction by adjusting the two-part model to cater to a specific problem: estimating high healthcare costs.

This research aims to explore two-part models to estimate costs and to assess whether these models perform better in estimating high healthcare costs for somatic care compared to the LS regression method. This study seeks to answer the following research question: ***How do two-part models perform relative to the least squares method in predicting high healthcare costs for somatic care?***

To address this research question, the research project is divided into three parts. The first part focuses on descriptive statistics and exploring thresholds. This part seeks to answer the following sub-question:

1. *How can high healthcare costs be defined and what are the key risk characteristics of individuals with high healthcare costs?*

The second part of the research focuses on predicting healthcare costs using two-part models and comparing these results to the least squares method. This part seeks to answer two sub-questions, which are:

2. *To what extent does a two-part model improve the outcome of the standard evaluation measures relative to the least squares method for somatic care?*
3. *To what extent does a two-part model reduce under- and overcompensation for high-cost individuals and chronic disease subgroups relative to the least squares method for somatic care?*

The third part of the research focuses on conceptually comparing two-part models to the Dutch risk equalisation model and discusses alternative estimation methods. This part seeks to answer the following sub-question:

4. *What are the pros and cons of two-part models and potential alternatives for predicting high healthcare costs?*

If concluded that a two-part model has better results in comparison to the LS regression method, the implementation of the two-part model into the equalisation scheme could be an interesting research topic for the WOR research program. This may result in a better prediction of healthcare costs, which results in a reduction of incentives for risk selection. Furthermore, with an accurate compensation for high-cost individuals, the ex-post measure in the risk equalisation system might not be necessary anymore.

The outline of this thesis is as follows. Chapter two of this thesis discusses the theory of regulated competition and risk equalisation. Furthermore, the theory on high-cost users, estimation methods and evaluation methods are discussed in this chapter. Chapter three describes the methodology, in which the data and data analyses for this thesis are set out. In this chapter, the ethics, validity and reliability of the research are also discussed. Chapter four presents the results of the data analyses. Lastly, chapter five gives a summary of the key findings, provides recommendations for future research, discusses policy implications and answers the research questions.

2. Theoretical framework

In this chapter, the theoretical framework supporting this study is discussed. For this research, it is important to have a good understanding of risk equalisation, insurance theory and statistical approaches. Therefore, firstly, the theory on regulated competition in the Dutch basic health insurance will be outlined. Next, the role of risk equalisation will be discussed. Thirdly, the characteristics of high-cost users and methods for identifying them will be presented. Fourthly, estimation methods for healthcare spending will be discussed. Finally, evaluation measures for estimation models will be presented. The literature was identified by reviewing key publications on the relevant literature topics of this research. Afterwards, additional literature was identified through the reference lists of these studies and by reviewing related studies from other authors.

2.1 Regulated competition in the Healthcare Insurance Act

2.1.1 Regulations in the Healthcare Insurance Act

The Dutch healthcare system is based on regulated competition, drawing on elements of Enthoven's model of managed competition (Enthoven, 1993). The regulations set out by the government are based on fairness principles. This means, for instance, that everyone in the Netherlands has the right to the same affordable healthcare regardless of their age, health state and welfare (Ministerie van Volksgezondheid, 2019). The Healthcare Insurance Act has multiple regulations set out, which ensure fairness principles in healthcare (*Zorgverzekeringswet*, 2026).

Firstly, there is a prohibition on risk-rated premiums within the basic health insurance. Risk rating is a process in which an insurer categorises its enrollees based on their individual risk. By charging a premium which covers the individual costs, the affordability of health insurance is at risk for people with high expected healthcare costs. For this reason, a standard community-rated premium is obligated. This means that all enrollees, despite their risk profile, pay the same standardised premium (Van Kleef et al., 2018). This community-rated premium results in predictable profits and losses per enrollee. These predictable profits and losses provide the rationale for risk equalisation, which will be further elaborated in Paragraph 2.2. Furthermore, the Dutch basic health insurance has a mandatory open enrolment. This obligates insurers to accept every person required to have insurance (Schut & Varkevisser, 2016).

2.1.2 Competition in the Healthcare Insurance Act

Insurers are allowed to selectively contract providers, which enables them to act as a countervailing power to healthcare suppliers (Douven, 2000). This power is exercised by critically contracting suppliers and negotiating on volumes, quality and price (Bes et al., 2012). By contracting selectively, insurers can attract enrollees, contain costs and ensure quality of care (Bes et al., 2016). Furthermore, insurers are partly allowed to compete on premiums. The nominal premium of an insurer is made up of two parts. Firstly, there is the standardised premium, which is the component that covers the expected healthcare costs. Secondly, there is the loading fee, which covers additional costs such as administrative expenses, solvency requirements and potential profit margins. Insurers can compete by offering lower premiums, which for example, can be achieved by maintaining a relatively low loading fee (Schut & Varkevisser, 2016). In addition, insurers compete on the service and quality of their policies. Policy quality refers to aspects such as the contracted provider network and the level of reimbursement for non-contracted providers. Service provision refers to waiting list mediation and claims processing (Netherlands Authority for Consumers and Markets, 2016). It should be noted that

insurers also supply supplementary insurance policies. These policies offer insurance for healthcare services which are not covered in the basic health insurance, which means that the regulations from the government do not apply to these policies (Rijksoverheid, 2016). Even though these policies do not fall under the Healthcare Insurance Act, insurers compete on these policies. This is because enrollees often purchase supplementary insurance from the same insurer that provides their basic insurance (Nederlandse Zorgautoriteit, 2015).

2.2 The role and relevance of risk equalisation

2.2.1 Rationale: avoiding risk selection and creating a level playing field

The risk equalisation system aims to create a level playing field among insurers. Insurers have enrolment portfolios that differ from one another. This results in some insurers having relatively many high-risk enrollees, while other insurers have a low-risk population. These different starting positions create an uneven playing field between insurers, which undermines competition. The risk equalisation system is applied to ensure a level playing field between insurers (Ministerie van Volksgezondheid, Welzijn en Sport, 2017).

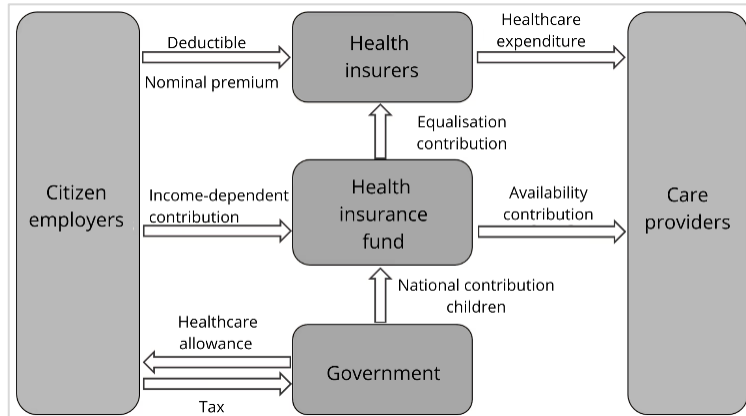
The second reason for implementing a risk equalisation system is to counteract incentives for risk selection. As a result of the prohibition on risk-rated premiums, there are predictable profits and losses among enrollees. These predictable profits and losses create incentives for insurers to engage in risk selection. This is because a competitive market leads to the equivalence principle. This principle states that an insurer seeks to minimise differences between premiums and expenditures. When risk-rated premiums are not allowed, insurers can achieve equivalence through risk selection (Van de Ven et al., 2013). The risk equalisation system counteracts these incentives for risk selection. Risk selection can occur through various channels. Insurers may design insurance policies that are less attractive to certain (patient) groups, with the aim of not enrolling individuals with a predictable loss (Van de Ven et al., 2013). For example, an insurer could selectively contract healthcare providers that are less preferred by chronically ill patients. Furthermore, insurers could offer supplementary insurance policies that only appeal to healthy individuals. Risk selection can also occur through marketing strategies by, for example, aiming the advertisement of insurance policies to healthy individuals. These selection activities lead to undesirable effects such as risk solidarity not being obtained. This occurs when undercompensated individuals end up having different insurance contracts than other individuals. Furthermore, selection activities may reduce the quality of care available to undercompensated individuals. This is because offering insurance policies with high-quality care for undercompensated individuals attracts these individuals. Finally, risk selection activities could undermine efficiency incentives for insurers when risk selection is more profitable than efficiency improvements (van Kleef et al., 2016).

2.2.2 The process of risk equalisation

Risk equalisation is applied within the Dutch basic health insurance (Netherlands Authority for Consumers and Markets, 2016). It is the process of financially compensating insurers who have a relatively large number of high-risk individuals in their risk portfolio. This system mostly ex-ante compensates insurers for predictable losses and profits among their enrollees (Van de Ven et al., 2013). This compensation is done through an equalisation contribution via the health insurance fund (Ministerie van Volksgezondheid, 2019). This fund is part of the funding flow for the Healthcare Insurance Act, which is shown in Figure 1. The contribution received by an insurer is based on the sum of the individual contributions of all enrollees. Risk equalisation in the Netherlands works via

cycles. The equalisation cycle starts one year prior to the equalisation year. This year, an estimate of the expected equalisation contribution is made. With this knowledge, health insurers can calculate their insurance premium. The cycle ends four years after the equalisation year. This is when the equalisation contribution is made definitive (Ministerie van Volksgezondheid, 2019).

Figure 1: Funding flow Healthcare Insurance Act



Source: Zorginstituut Nederland (2019), Translated

2.2.3 The Dutch risk equalisation system

Risk equalisation in the Dutch healthcare system is done by applying three models. The first model focuses on somatic healthcare and the second model focuses on mental healthcare. Both models use risk characteristics as variables which help estimate the healthcare costs for these specific types of healthcare. Per enrollee, a statistical model calculates the expected costs in the equalisation year according to the individual's risk characteristics (Eijkenaar & Van Vliet, 2017). Risk characteristics are characteristics of an individual which are associated with healthcare costs, such as age and sex (Ministerie van Volksgezondheid, Welzijn en Sport, 2019). Risk adjusters that are used in the 2026 risk equalisation model for somatic care are shown in Table 1, including the number of risk classes and a brief description. As shown in Table 1, the 2026 risk equalisation model includes a 'seasonal worker' risk adjuster. This adjuster is not used in this thesis because the results of this research apply only to the population covered by the Healthcare Insurance Act, excluding non-residents. For more information on the risk classes for each risk adjuster, see Ministerie van Volksgezondheid, Welzijn en Sport (2025b).

Table 1: Risk adjusters somatic model 2026

Risk adjuster (abbreviation)	Risk classes	Description
Age/sex	42	Risk classes based on an interaction between age and sex
Pharmacy cost groups (PCG)	51	Morbidity indicator derived from the use of specific pharmaceuticals in the prior year
Diagnostic cost groups (DCG)	27	Morbidity indicator derived from specific hospital treatments in the prior year
Source of income, interacted with age (SOI)	36	Socioeconomic indicator based on the source of income (e.g., employed, self-employed, social security...)
Region	10	Clusters of ZIP codes based on regional characteristics
Socioeconomic Status, interacted with age (SES)	12	Socioeconomic indicator based on the mean per-person household income
Persons per address, interacted with age (PPA)	19	Morbidity indicator based on whether the individual lives in a care facility and whether the person lives alone
Multiple-year high costs (MHC)	9	Morbidity indicator based on whether the individual has had high healthcare costs in the three previous years
Physical therapy diagnosis groups (PTD)	5	Morbidity indicator based on whether the individual received physical therapy for a chronic disease in the prior year
Multiple year Costs Home Care (MCHC)	10	Morbidity indicator based on whether the individual has had costs for home care in the previous three years
Historical somatic morbidity (HSM)	2	Morbidity indicator based on whether the individual was in a positive somatic morbidity class three years prior
Indicator pregnancy (IP)	4	Morbidity indicator based on whether the individual was pregnant and/or gave birth in the equalisation year.
Historical cost for care aids (HCA)	7	Morbidity indicator based on whether the individual had costs for care aids in the previous three years in the top 0.1%, 0.2%, 0.5%, 1%, 3% or 9%
Missing costs history (MCH)	6	Morbidity indicator based on whether the individual was (partly) not insured in the last, second-last and third-last year.
Seasonal worker (SW)	3	An indicator based on the inhabitant status of the enrollee, shows whether the enrollee is a seasonal worker or resides abroad.
Total number of risk classes	243	

Source: Ministerie van Volksgezondheid, Welzijn en Sport (2025b)

This thesis focuses on the somatic risk equalisation model. Risk adjusters that are used in the 2026 risk equalisation model for mental healthcare are shown in Appendix 1, including the number of risk classes and a brief description. In addition to the somatic and mental healthcare model, there is the deductible model. This model accounts for the costs that must be paid out-of-pocket by the enrollee. The somatic and mental-healthcare equalisation models do not account for this out-of-pocket payment (Gupta Strategists, 2025a). The deductible model estimates the extent to which an individual reaches their deductible. Risk adjusters that are used in the 2026 risk equalisation model for the deductible are also shown in Appendix 1. The total estimated costs under the Healthcare Insurance Act are calculated by adding the predicted costs from the somatic risk equalisation model to the predicted costs from the mental healthcare model, and then subtracting the predicted out-of-pocket payment from the deductible model. Risk equalisation therefore only applies to healthcare costs that are paid by insurers under the Healthcare Insurance Act.

2.2.4 International context

Risk equalisation is applied in several other countries. Firstly, risk equalisation is applied in Germany for the German sickness funds. A study on the German model has shown that this model has a systematic undercompensation for multi-morbidity (Buchner et al., 2013). Secondly, risk equalisation

is applied in Switzerland by morbidity-based health plan payments. Research by Kauer et al. (2020) shows that the Swiss system has extreme under- and overcompensation for the top and bottom percentiles. Thirdly, risk equalisation is applied in Australia within the private health insurance. Here, the system is based only on risk sharing (Henriquez et al., 2023). In addition to these countries, risk adjustment is also applied in other countries, such as Belgium (Schokkaert & Van de Voorde, 2000), Israel (Shmueli et al., 2003), the United States (Field & Jette, 2007), Chile (Vargas & Wasem, 2006) and Ireland (Armstrong, 2018).¹ The challenges outlined in this paragraph show that a perfect equalisation system does not yet exist.

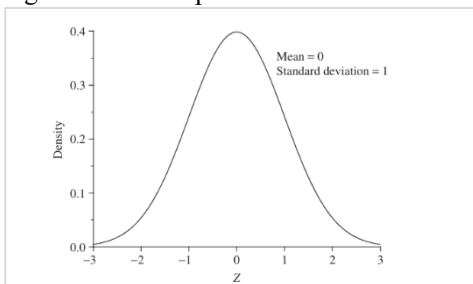
2.3 Methods for estimating healthcare spending

The literature defines multiple methods to estimate healthcare spending. In the upcoming paragraphs, methods to estimate healthcare spending are outlined.

2.3.1 Least squares regression

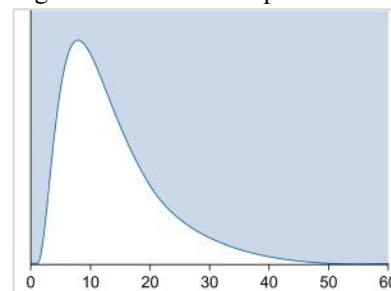
Healthcare costs can be estimated using least squares regression. In broad terms, the Dutch equalisation model resembles a least squares (LS) regression model. An LS regression is the most widely used regression and is bound to the normality assumption (Tong, 2025). This assumes that, with a large sample size, the distribution has the least amount of variance (Schmidt & Finan, 2018). Because of the normality assumption, the distribution of the data is assumed to be bell-shaped (Smith, 2015). This bell-shaped distribution is shown in Figure 4. Although the equalisation model uses data on all people with basic health insurance, which suggests a large sample size, the normality assumption does not hold. Healthcare expenditure data is skewed because of individuals with high healthcare costs. This results in the distribution of healthcare expenditure taking the shape of Figure 5: a bell-shape with a long right-tail. This skewed distribution can create problems in estimating healthcare spending using an LS regression (Tong, 2025).

Figure 4: Bell-shaped distribution



Source: Smith (2015)

Figure 5: Healthcare expenditure distribution



Source: Siegel (2012)

2.3.2 Machine learning

Healthcare costs can be estimated using machine learning. Rose (2016) compared the performance of statistical approaches to a machine learning framework for risk adjustment formulas. This study concluded that simplifying formulas reduces insurers' ability to manipulate risk adjustment mechanisms (Rose, 2016). Within the Dutch risk equalisation context, Ismail et al. (2024) studied the application of machine learning models to estimate healthcare spending. The application of machine learning explained an additional 1.5% of medical expenditure. In this study, it was concluded that the

¹There are forms of collaboration between the countries which apply a risk equalisation system. An example of this is the Risk Adjustment Network, which is a joint effort between health economists. This network exchanges scientific findings in the field of risk equalisation systems (*Risk Adjustment Network*, n.d.).

application of machine learning improves the estimation of the healthcare costs of high- and low-cost subgroups. However, machine learning has not yet been applied in the risk equalisation system. This may be explained by the fact that, even though these algorithms improve the predictive performance of the model, machine learning reduces the transparency of the risk equalisation system (Ismail et al., 2024).

2.3.3 Constrained regression

Healthcare costs can be estimated using constrained regression. Reitsma et al. (2025) researched the use of constrained regression in Medicare. This study showed that constrained regression introduces fairness objectives with minimal impact on overall fit. The application of constrained regression reduces under- or overcompensation in risk equalisation. This can be done by subjecting the goal of the model to a maximum under- or overcompensation (Van Kleef et al., 2017). Since the Dutch 2024 risk equalisation model, constrained regression has been incorporated into the somatic model (Minister van Volksgezondheid, Welzijn en Sport, 2023). In the equalisation model, constrained regression is applied to the Multiple-year Pharmacy Costs variable (Gupta Strategists, 2025a).

2.3.4 Two- and three-part model

Healthcare spending can be estimated using two- and three-part models. A two-part model models a conditional outcome as separate processes. This can be used for modelling healthcare costs (Belotti et al., 2015). Srivastava et al. (2021) researched the use of a two-part model on a skewed expenditure distribution. In this study, first, the probability of incurring expenditure was modelled per household. Secondly, the amount of expenditure for the households was estimated using an LS regression. Karlsson et al. (2024) implemented a three-part model for healthcare expenditure distributions. In this study, the researchers extended the two-part model by creating a decision within the non-zero category, using a threshold condition. This extension was done to provide a more nuanced insight into healthcare spending patterns.

This thesis aims to take a different approach by modelling a two-part model with a non-zero threshold condition, inspired by the three-part model from Karlsson et al. This approach is applied because the zero-mass problem (as described by Karlsson et al. (2024)) is not as critical for somatic care because individuals with no healthcare costs are not as prevalent. A decision within the non-zero category is important because of the skewed expenditure distribution.

This model consists of three key variables. Equation 1 defines variable Y_1 which shows the healthcare costs that are below threshold T .

$$Y_1 = Y \mid Y < T \quad (\text{Equation 1})$$

Equation 2 defines variable P , which shows the probability of an individual to incur high healthcare costs. The value of P lies between zero and one, where a value close to one indicates a high probability of incurring high healthcare costs.

$$P = Pr(\text{Costs} > T) \quad (\text{Equation 2})$$

Equation 3 defines variable Y_2 which shows the healthcare costs that are above threshold T .

$$Y_2 = Y \mid Y > T \quad (\text{Equation 3})$$

The core of the two-part model is that the three components are estimated separately, after which these values are combined to predict the total healthcare costs. When applying a two-part model, there are several decisions which must be made to complete the model. Firstly, the threshold which

determines high healthcare costs must be defined. Secondly, the type of regression for each part of the model has to be specified. Within the two-part model approach taken in this thesis, a regression is first used to determine the probability of incurring high healthcare costs. Afterwards, regressions are used to estimate the healthcare costs below and above the threshold.

2.3.5 Stepwise regression

The use of stepwise regression can give insight into which variables have a significant impact on the equalisation model. Stepwise regression is a type of hierarchical regression which determines which variables end up in the model (Miller et al., 2022). There are three types of stepwise regression. Firstly, there is forward stepwise regression, which adds variables until no more variables can significantly improve the model's fit. Downsides of this method include the risk of increased model complexity and the risk of omitted variable bias. Secondly, there is backwards stepwise regression. In this approach, variables are removed that do not significantly contribute to the model. This method has less risk of becoming too complex. Lastly, there is bidirectional stepwise regression. This type of regression considers adding and removing variables at each step (Huang et al., 2025). Stepwise regression can be incorporated into the equalisation model by applying stepwise regression to the regressions used in the model.

2.4 High-cost users

A small share of the population has high healthcare costs, while the majority of the population incurs low healthcare costs (Karunayawong et al., 2024). The literature defines individuals with high healthcare costs as high-cost users. In the upcoming paragraphs, methods to define high-cost users are outlined.

2.4.1 High-cost users according to an absolute monetary threshold

High-cost users can be identified by an absolute monetary threshold. To set this threshold, it is important to have information on healthcare spending in the country of interest, which in this case is the Netherlands. In 2024, the total healthcare spending specifically for the Healthcare Insurance Act was €57.2 billion (Statistics Netherlands, 2025a). This comes down to an average spending of €3,286 per person in 2024 (Statistics Netherlands, 2025b). An absolute monetary threshold is applied within the ex-post compensation for high costs in the 2026 risk equalisation model. This measure is applied to help compensate insurers for extremely high costs (Hamstra et al., 2023). The research on this ex-post measure defined an absolute monetary threshold, which states that, on average, the costs per insured must be €100,000 to qualify for the high-cost group.² According to this research, another precondition is that the high-cost individual should be undercompensated for €50,000 minimum (Hamstra et al., 2023). A downside of this method is that an absolute monetary threshold does not take the expenditure distribution and possible cost increases in time into account. This represents a weakness of this threshold method.

2.4.2 High-cost users according to percentiles

High-cost users can be identified using percentiles of the healthcare expenditure distribution, where individuals whose expenditure exceeds a set percentile threshold are classified as high-cost users. An example of such a threshold is shown in the study by Jansen et al. (2025), which shows that the costs for the 99th percentile of geriatric rehabilitation users in the Netherlands in 2023 were €46,830. A

² This amount (€100,000) is not used as a threshold for the ex-post compensation in the equalisation model. Because of other reasons, €400,000 is used as a threshold for ex-post compensation for high costs (Hamstra et al., 2023).

study by Dziegielewska et al. (2021) applied this method by defining high-cost users as individuals in the upper tenth percentile of the total healthcare costs. This study showed that, on average, high-cost users were 64.2 years old and had more comorbidities (Dziegielewska et al., 2021). A downside of this method is that it focuses on the distribution of costs rather than the distribution of individuals, and it is therefore sensitive to outliers.

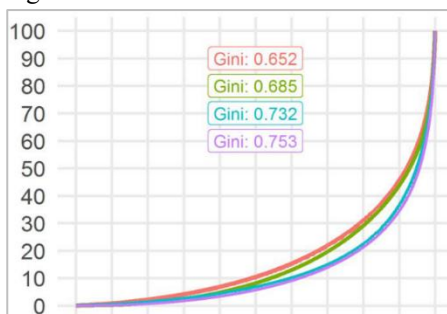
2.4.3 High-cost users according to top-percentages

High-cost users can be defined as the top x-percentage of individuals with the highest healthcare expenditures. This method ranks individuals based on their expenditures and selects a proportion of the population. Within the WOR-assessment framework for research on risk equalisation, the measurement suggestion list provides recommended top-spender percentages for high-cost individuals. These suggestions consist of the 15% of individuals with the highest costs and individuals with the 1%, 3% and 5% highest costs in the previous year (Ministerie van Volksgezondheid, Welzijn en Sport 2024). A study by Wammes et al. (2018) used this method and defined high-cost users as the top 5% and top 10% of individuals with the highest costs. This study showed that high-cost users have a high prevalence of having multiple conditions. Another study by Wammes et al. (2017) specifically looked at the characteristics of high-cost users in the Netherlands in 2013. In this study, high-cost users were identified as two groups: the top 1% and the top 2% - 5% beneficiaries of the total costs. This research showed that 76% of the costs of the top 1% spenders were made up of hospital costs. McGuire et al. (2021) also implemented this approach. In this study, extremely high residual spenders were researched by including the top 1% and top 0.1% individuals with the highest residuals. This study concluded that individuals with high residual spending account for a large share of unexplained variance (McGuire et al., 2021).

2.4.4 High-cost users according to the Lorenz curve

High-cost users can be identified according to the Lorenz curve. A Lorenz curve is a graphical tool that shows inequalities within a population (Mauguen & Begg, 2016). The Lorenz curve can be used to identify high-cost users by determining the point at which the healthcare costs become highly unequal. Espinosa et al. (2025) applied a Lorenz curve to gain insight into the inequality in healthcare costs in Colombia. This was done by plotting multiple Lorenz curves, which show the cumulative percentage of expenditures against the cumulative percentage of population. This study showed that in Colombia, less than 10% of the population accounts for over 50% of the expenditure. An example of a Lorenz curve based on healthcare expenditure is shown in Figure 2.

Figure 2: Lorenz curve of the healthcare expenditure for the Colombian population



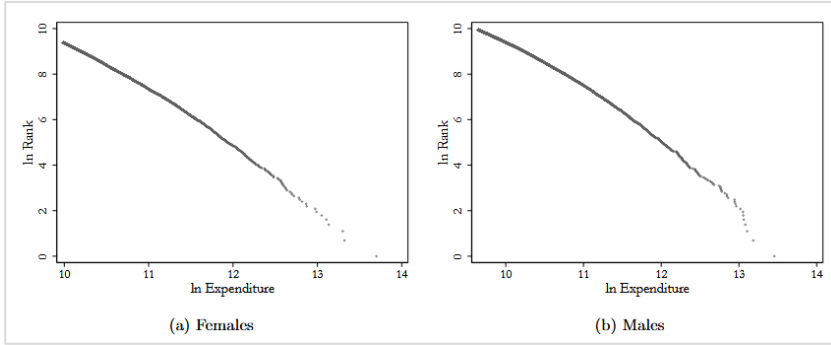
Source: Espinosa et al. (2025)

2.4.5 High-cost users according to the Pareto-tailed distribution

High-cost users could be identified according to the Pareto tail method. This method relies on the fact that the distribution of healthcare costs has a heavy right tail that can be modelled using a Pareto-

tailed distribution (Karlsson et al., 2024). This method is applied by Karlsson et al. (2024) to set the threshold in the non-zero category of a three-part model. This approach is applied by firstly sorting the healthcare costs from high to low and log-transforming these values. Afterwards, these values are set of against the log-transformed rank of each individual, based on the healthcare costs. An example of this plot is shown in Figure 3. In this method, the plot suggests a linear relationship when the data points align along a straight line. The point in the plot at which the observations start suggesting a linear relationship can be applied as a threshold for high healthcare costs.

Figure 3: Plots of Natural Logarithms of Rank against Expenditures



Source: Karlsson et al. (2024)

2.5 Evaluation measures and the WOR-assessment framework

2.5.1 The WOR-assessment framework

To evaluate risk equalisation models, the WOR-assessment framework has measures outlined. This framework is developed by the Ministry of Health, Welfare and Sport and concerns the criteria of risk equalisation: payment fit, efficiency, manageable complexity, validity, measurability and stability (Ministerie van Volksgezondheid, Welzijn en Sport, 2024).

Payment fit focuses on the degree of fit between costs and reimbursements at the individual- and subgroup level (Ministerie van Volksgezondheid, Welzijn en Sport, 2024). A list of the standard evaluation subgroups for the payment fit, according to the framework, is taken up in Appendix 2. For each subgroup, the number of individuals in that group is determined, along with the Mean Financial Result of the equalisation model for that subgroup. The Mean Financial Result can be calculated using Equation 4, where ‘g’ refers to a specific subgroup.

$$\text{Mean Financial Result}_g = \frac{\sum_{i=1}^g (\hat{Y}_i - Y_i)}{n_g} \quad (\text{Equation 4})$$

In addition to the subgroups, statistical measures are outlined in the framework. Firstly, this is the R-squared measure, which measures what proportion of the variation in healthcare costs is explained by the risk characteristics (Chicco et al., 2021). The equation for this measure is shown in Equation 5. Secondly, there is the Cummings Prediction Measure (CPM), which captures the proportion of the variation in absolute differences in healthcare costs that is explained by a risk equalisation model (Beck et al., 2020). The equation for this measure is shown in Equation 6.

$$R^2 = 1 - \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (\text{Equation 5})$$

$$\text{CPM} = 1 - \frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{\sum_{i=1}^n |Y_i - \bar{Y}|} \quad (\text{Equation 6})$$

The second criterion of risk equalisation is efficiency. This criterion entails that the equalisation model should not discourage working efficiently. In the WOR framework, this criterion is set out in sub-criteria: cost-containment incentives, volume-containment incentives and incentives for appropriate care. Cost-containment incentives can be assessed by evaluating how much money is allocated based on characteristics directly linked to costs. Volume-containment incentives can be assessed by evaluating how much money is allocated based on characteristics linked to utilisation. Incentives for appropriate care can be evaluated by medical specialists, who help determine what appropriate care actually is (Ministerie van Volksgezondheid, Welzijn en Sport, 2024).

The third criterion of risk equalisation is manageable complexity. This criterion entails that the risk equalisation model should be manageable in implementation and in research. To assess the models' complexity, the WOR-assessment framework has a list of possibly complex process steps which have to be discussed. The steps are: the collection of source data, preparation of execution files, characteristic derivation, interpretation of results, application of ex-post measures, calculation of payment weights and contribution determination. The discussion of these process steps should be done by the researcher. However, other parties, such as the Dutch Healthcare Institute should also be involved in this discussion, as they are the responsible party for the execution of the model and therefore experts on complexity (Ministerie van Volksgezondheid, Welzijn en Sport, 2024).

The final criteria of risk equalisation are validity, measurability and stability. Firstly, validity is obtained when there is a systematic relationship between the model change and/or characteristic and the predicted healthcare costs. To assess validity, desk research on the new characteristic needs to be done. Furthermore, measurability is obtained when a characteristic classifies insured persons into unambiguous classes objectively and reliably. This is a knock-out criterion. Stability is obtained when equalisation criteria are representative for later years. To assess stability, sensitivity analyses can be done (Ministerie van Volksgezondheid, Welzijn en Sport, 2024).

2.5.2 Beyond the standard WOR-assessment framework

The WOR-assessment framework includes a list of suggested subgroups in the appendices. One of these subgroups contains enrollees who, in 2020, belonged to the 15% of people with the highest somatic healthcare costs. This subgroup could be interesting to apply in addition to the standard subgroups because it specifically focuses on individuals with high healthcare costs. The Mean Absolute Error (MAE) measure is not mentioned in the framework. However, this measure can also be applied to evaluate the equalisation models. The MAE measures how accurate the prediction of a model is by showing the average amount of units the model's predictions are off from reality (Reich et al., 2016). The equation for this measure is shown in Equation 7.

$$\text{Mean Absolute Error} = (1/n) \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (\text{Equation 7})$$

2.5.3 The context of the WOR-assessment framework

The WOR-assessment framework is used to evaluate models which have been researched according to the WOR research program. This workgroup is led by the Ministry of Health, Welfare and Sport and aims to continuously improve the Dutch risk equalisation model. Each year, stakeholders such as health insurers, the Dutch Healthcare Institute and researchers determine the research agenda of the WOR. The results of these studies are translated into recommendations to the Minister of Health, Welfare and Sport. The minister then determines which adjustments will be implemented into the equalisation model. The updated equalisation model is presented annually in October (Ministerie van Volksgezondheid, Welzijn en Sport, 2017). Research on alternative estimation methods is not

included in the 2025-2026 research agenda of the WOR (Werkgroep Ontwikkeling Risicoverevening, n.d.). However, the models in this study are evaluated according to the WOR-assessment framework to align this thesis with the WOR research and policy framework. If the results from this thesis indicate that two-part models perform better in estimating high healthcare costs than LS regression, this topic could be interesting for the upcoming WOR research agenda.

3. Methodology

The study design of this research is a quantitative simulation on microdata. This chapter provides an overview of the methodology used to estimate and compare risk equalisation models. In this chapter, the available data and the analysis processes are described. Furthermore, the evaluation methods for the models are discussed. Lastly, the ethics, validity and reliability of the research are discussed.

3.1 Data

The data for this study was made available through the Risk Equalisation Research Program, an initiative of the Ministry of Health, Welfare and Sport and the Association of Health Insurers. The dataset contained information on Dutch individuals with mandatory basic health insurance in 2023. The variables in this dataset correspond to the variables outlined in the theoretical framework (Table 1, Paragraph 2.2.3, *excluding the SW risk adjuster*). This data allowed for the replication of the 2026 risk equalisation model. In addition to the risk adjusters, additional variables were made available. These variables, for example, allowed for evaluation of the models on a subgroup level. An overview and description of all available additional variables is shown in Appendix 3. The dataset was accessed via the microdata portal from the Dutch Central Bureau for Statistics. Permission to use the data had been obtained via the Risk Equalisation Research program from the Erasmus University Rotterdam, faculty Erasmus School of Health, Policy and Management.

Table 2 presents the descriptive statistics for all individuals in the dataset. The mean healthcare costs per individual in 2023 were €2,834. The population mostly consists of individuals who are between the ages of 18 and 59. 12% of the population has had specific hospital treatments in the prior year. Furthermore, 26.8% of the population has used specific pharmaceuticals in the prior year. The population is equally distributed between the sexes. Furthermore, the difference between the number of insured individuals and the number of insured years can, for example, be attributed to the death of individuals within the insurance year.

Table 2: Descriptive statistics

Mean healthcare costs	€2,834
% Age 0-4	5.3%
% Age 5-17	13.6%
% Age 18-39	27.7%
% Age 40-59	25.8%
% Age 60-74	17.8%
% Age 75+	9.8%
% at least one DCG	12.0%
% at least one PCG	26.8%
% men	49.6%
% women	50.4%
Number of insured individuals	17,736,376
Number of insured years	17,400,000

Based on the cost for the year 2023, and 17,736,376 observations. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

3.2 Data analysis

The data analysis applied in the research consisted of four steps. Each of these steps is discussed in a separate paragraph. Throughout the analyses, healthcare spending was annualised, and individuals were weighted according to the fraction of the year for which they were insured. An example of this is an individual who was insured for three months of the year and had healthcare costs of €1,000. For

this individual, annualised spending equals €4,000 ($1,000/(3/12)$), and this individual was given a weight of 0.25 ($3/12$). By annualising spending and weighing individuals, the mean healthcare costs, mean age and mean under/overcompensation can be interpreted as the mean of these variables per insured year.

3.2.1 Descriptive statistics

Firstly, descriptive statistics were used to gain insight into the range and skewness of the data. The skewness of the data had been derived via Stata's skewness value. Furthermore, a Lorenz curve was plotted, which showed the inequality of healthcare costs across the population.

3.2.2 Simulating the 2026 risk equalisation model for somatic care

The Dutch 2026 risk equalisation model was simulated using LS regression. However, the 2026 equalisation model includes restrictions. These restrictions were not modelled to improve comparability between the simulated equalisation model and the two-part models. This choice will be further discussed in Paragraph 3.4. To simulate the equalisation model, the risk characteristics from the 2026 equalisation model were used as independent variables in the LS model. A regression was used to estimate the beta coefficients. The relationship between these coefficients and the independent variables is shown in Equation 8. The estimated beta coefficients were used to predict the healthcare costs.

$$Y_{ls} = \beta_0 + \beta_1 X_1 + \dots + \beta_{240} X_{240} + e \quad (\text{Equation 8})$$

3.2.3 Constructing the two-part models

To construct the two-part models, the required variables were defined in the dataset. Next, the necessary regression models were specified using these variables. Finally, healthcare costs were estimated by applying the predictive models derived from these regressions. These different steps are further elaborated in the paragraphs below.

Variable construction

Threshold T , which defines high healthcare costs, was specified as a monetary value. As no standardised threshold exists, three alternative values for T were applied. These values were derived from the theory discussed in Paragraph 2.4. To assess which threshold is most appropriate, the thresholds were applied in separate two-part models. Variable P was created to indicate whether an individual exceeded threshold T , based on the actual healthcare costs in 2023. If an individual exceeded the threshold, that individual got value 1 for variable P ; otherwise, the individual got value 0. Variable P is further defined in Equation 9.

$$P = \begin{cases} 1 & \text{if } costs > T \\ 0 & \text{if } costs \leq T \end{cases} \quad (\text{Equation 9})$$

Next, variable Y_1 was created by cutting of individual-level costs at the threshold. Variable Y_1 is defined in Equation 10.

$$Y_1 = \min(Costs, T) \quad (\text{Equation 10})$$

Variable Y_2 was created by subtracting the threshold amount from the costs and setting the minimum value to zero. Variable Y_2 is defined in Equation 11.

$$Y_2 = \max(Costs - T, 0) \quad (\text{Equation 11})$$

Regression model specification

Descriptive statistics were used to describe the characteristics of individuals who exceeded the threshold and thus had a value of 1 for P. These statistics included the frequency of people with high healthcare costs, which was sorted based on age. Furthermore, the average costs of people with high healthcare costs were reported, as well as the interquartile range and median healthcare costs. Lastly, key risk characteristics of individuals with high healthcare costs were reported. Afterwards, a logistic regression was modelled using variable P and all risk characteristics as x-variables to estimate the probability of an individual incurring costs above the threshold. This is shown in Equation 12.

$$P = \frac{1}{1+e^{-(\beta_0+\beta_1X_1+\dots+\beta_{240}X_{240})}} + e \quad (\text{Equation 12})$$

By estimating an LS regression, using variable Y_1 and all risk characteristics as x-variables, the expected costs below the threshold were determined. The equation for the expected costs below the threshold is shown in Equation 13.

$$Y_1 = \beta_0 + \beta_1X_1 + \dots + \beta_{240}X_{240} + e \quad (\text{Equation 13})$$

By estimating an LS regression for variable Y_2 , using all risk characteristics as x-variables, the expected costs above the threshold were determined. The equation for the expected costs above the threshold is shown in Equation 14.

$$Y_2 = \beta_0 + \beta_1X_1 + \dots + \beta_{240}X_{240} + e \quad (\text{Equation 14})$$

Prediction models

To calculate the total expected healthcare costs, Equations 15, 16, and 17 were applied. Equation 15 shows the predicted individual healthcare costs below threshold T. Equation 16 shows the predicted individual healthcare costs above threshold T. Equation 17 shows the predicted individual probability of incurring healthcare costs above threshold T. When calculating this probability, there is a risk of perfect multicollinearity, which results in several observations being omitted.³ To ensure that all observations are used in the model, the variables which exhibit perfect multicollinearity were not used in Equation 17.

$$\hat{Y}_{1i} = b_0 + b_1X_{1i} + \dots + b_{240}X_{240i} \quad (\text{Equation 15})$$

$$\hat{Y}_{2i} = b_0 + b_1X_{1i} + \dots + b_{240}X_{240i} \quad (\text{Equation 16})$$

$$\hat{p}_i = \frac{1}{1+e^{-(b_0+b_1+\dots+b_k)}} \quad (\text{Equation 17})$$

The equation for the total individual expected healthcare costs, noted as $\hat{Y}_{3i}$, is shown in Equation 18.

$$\hat{Y}_{3i} = \hat{Y}_{1i} + \hat{p}_i \cdot \hat{Y}_{2i} \quad (\text{Equation 18})$$

Cost transformation

Two-part models do not use LS estimation logic. As a result, the mean predicted healthcare costs from the two-part models did not automatically align with the mean predicted healthcare costs of the LS regression model or with the observed mean healthcare costs. To be able to compare the predicted healthcare costs across the two-part models, the LS regression model and the observed costs, the two-part model predictions were transformed. This was done by, for each two-part model, calculating a

³ Perfect multicollinearity occurs when a variable only includes individuals who score a '1' on variable P and also score a '1' on the variable itself, causing the probability value to always equal '1'.

linear transformation factor equal to the ratio of the mean observed healthcare costs to the mean predicted healthcare costs. Subsequently, the individual predicted healthcare costs ($\hat{Y}_{3i}$) were multiplied by this factor. This transformation ensured that the mean predicted healthcare costs were aligned across models, which improved comparability.

3.3 Evaluation and additional analyses

3.3.1 The WOR-assessment framework and additional measures

The WOR-assessment framework was applied to determine if a two-part model improves the predictive performance of the equalisation model and meets the criteria. In total, three two-part models and one LS model were assessed and compared. The first evaluation was based on the payment fit. All models were evaluated using multiple subgroups. Firstly, this was the subgroup of enrollees with the presence of at least one chronic condition. Secondly, the subgroup of enrollees who have had pharmacy costs for multiple years was evaluated. Furthermore, the subgroup of enrollees who selected the maximum deductible and the subgroup of enrollees holding insurance contracts with restrictive conditions were also evaluated. Lastly, the subgroups of enrollees who belong to the 10% of individuals with the highest and lowest premiums were evaluated. For each subgroup, the Mean Financial Result was calculated using Equation 4 (Paragraph 2.5.1). Furthermore, the R-squared measure (Equation 5, Paragraph 2.5.1) and Cummings Prediction Measure (Equation 6, Paragraph 2.5.1) were applied to each model.

The complexity of each model was evaluated using the framework's list of complex elements. The evaluation criteria on validity, measurability, stability and efficiency were discussed by comparing the criteria outcomes of the two-part models to the criteria outcomes of the LS model.

Building on the evaluation based on the WOR-assessment framework, the suggested subgroup, which contains enrollees who, in 2020, belonged to the 15% of people with the highest somatic healthcare costs, was evaluated according to the method described above. This ensured specific evaluation for the high-cost subgroup. Furthermore, the Mean Absolute Error measure (Equation 7, Paragraph 2.5.2) was used to evaluate each model. This measure showed the average absolute amount of euros each model's prediction was off, which helped gain insight into the overall payment fit of the models.

3.3.2 Stepwise regression analysis

Comparability between the two-part models and the simulated equalisation model was improved by including all risk characteristics as x-variables in the estimated regressions. To gain insight into which variables have a significant impact on the two-part models, one two-part model was re-estimated using stepwise regression. The two-part model selected for this analysis was the most promising two-part model, based on the payment fit for high-cost individuals. Backwards stepwise regression was applied because this method reduces issues of model complexity (Huang et al., 2025). To determine which variables have a significant impact on the model, the p-value criterion was set to 0.01, as this is a commonly accepted p-value in the literature. After selecting the two-part model, a stepwise logistic regression was estimated for variable P and all risk characteristics as x-variables. This was done equivalent to Equation 17. Next, stepwise LS regressions were estimated for variables Y_1 and Y_2 . This was done equivalent to Equations 15 and 16. The outcome of this analysis showed the non-significant variables for each regression.

3.3.3 The two-part model compared to the 2026 risk equalisation model

The performance of the most promising two-part model, identical to the model used in the stepwise regression analysis, was compared to the 2026 risk equalisation model and the LS regression model. This comparison was limited to the payment fit of the models and gave insight into the performance of the two-part model relative to the 2026 model. Furthermore, this analysis provided an indication of the extent to which a two-part model could serve as an alternative to constrained regression.

3.4 Ethics, validity and reliability

The data used in this research contained health information which was pseudonymized to minimise ethical risks. Furthermore, the analyses were conducted within the secure environment of the data provider. The pseudonymization and secure environment limited the risk of harm to the individuals in the dataset.

The internal validity of this research depended on the extent to which the estimated models differed from, or aligned with, the 2026 risk equalisation model. The analyses were built upon the Dutch 2026 risk equalisation model and utilised real-world data from this model, which enhanced the internal validity of this research. Moreover, the use of full population data eliminated sampling bias. However, this research did not assess the stability of the results across different data years. To do so, the analyses would need to be repeated using datasets from different years. This may affect the internal validity of the research, as the results have not yet been proven to be stable. The external validity of this research was supported by the generalizability of broader patterns to other countries. Although specific results offered limited insights to other countries, broader patterns may lead to recommendations on repeating the analysis with data from other countries.

The reliability is supported by using the statistical software Stata. All data analysis steps were performed using StataNow 19.5 MP, with all commands executed through a do-file. This allows for replication of the analyses by other researchers, with access to the same data.

Several restrictions included in the 2026 Dutch risk equalisation model were omitted from both the LS regression model and the two-part models. These restrictions included: (1) the constrained regression based on the Multiple Year Pharma Costs (MPC) variable, which ensures an average Mean Financial Result of €0 for individuals who have had pharmacy costs for multiple years; (2) the age/sex risk adjuster classification, which ensures that payment weights for the age/sex risk adjuster sum to the total healthcare costs at the macro level; (3) the risk-class classification, which ensures that the payment weights for all risk characteristics other than age/sex sum to zero at the macro level. The implications of omitting these restrictions for the interpretation of the results will be discussed in the Discussion chapter of this thesis.

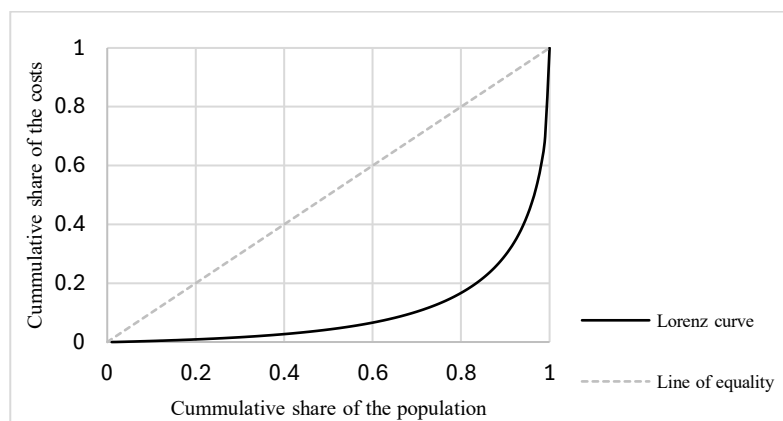
4. Results

In this chapter, the results from the analyses are discussed. The chapter starts by discussing the descriptive statistics of the data. Next, the results from the simulated LS regression model are presented. Afterwards, the thresholds used in the two-part models are discussed, as well as the characteristics of high-cost users. Lastly, the results from the evaluation and additional analyses are presented. These results consist of the WOR-assessment framework, the stepwise regression analysis and the 2026 model comparison.

4.1 Descriptive statistics

The distribution of the actual healthcare costs in 2023 is highly skewed. This is indicated by Stata's skewness statistic, which has a value of 44; values above 1 reflect a skewed distribution. This means that there are many individuals with low healthcare costs and a small number of individuals with high healthcare costs. Figure 6 presents the Lorenz curve of the cost distribution. In this figure, the cumulative share of the population is shown on the x-axis and the cumulative share of the healthcare costs is shown on the y-axis. The curve indicates that a small proportion of the population accounts for a large proportion of the healthcare costs. Approximately 20% of the population accounts for over 80% of the healthcare expenditure. This further confirms that the distribution of the costs is skewed.

Figure 6: Lorenz curve



Based on the cost for the year 2023, 17,736,376 observations and 100 bins. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

4.2 The LS regression model

Table 3 shows the summarised results from the LS regression model. The results from this model are compared to the actual healthcare costs and the results from the 2026 risk equalisation model. Table 3 shows that the standard deviation of the LS regression model is comparable to the standard deviation from the 2026 risk equalisation model. However, Table 3 also indicates that the standard deviation from these two models differs from the standard deviation of the actual healthcare costs. This indicates that neither the 2026 equalisation model nor the LS regression model gives a perfect prediction in terms of healthcare costs. However, this is rather logical given that a large share of healthcare costs remains unpredictable.

Table 3: Mean and standard deviation of actual healthcare spending and predicted healthcare spending according to the LS model

	Mean	Std. dev.
Predicted healthcare costs LS model	€2,834	5,643
Predicted healthcare costs 2026 model	€2,834	5,712
Actual healthcare costs	€2,834	10,077

Based on the cost for the year 2023, and 17,736,376 observations. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

4.3 The two-part models

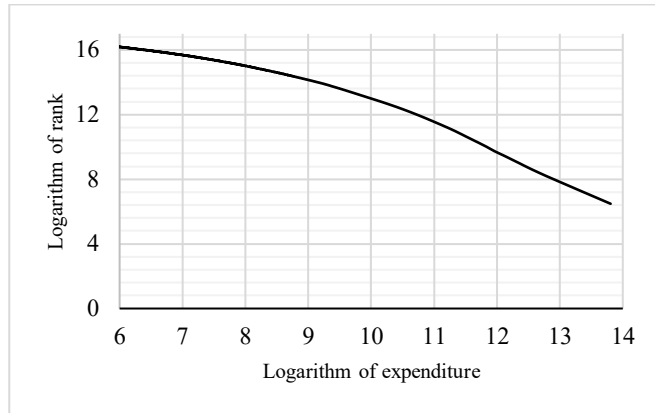
4.3.1 Constructing the two-part models

Threshold setting

The first two-part model defines high-cost users as the 5% of individuals with the highest healthcare costs. This threshold stems from the top-percentage approach (which was discussed in Paragraph 2.4.3) and aligns with the methodology of the Ministry of Health, Welfare and Sport. Furthermore, this threshold is consistent with previous studies (Wammes et al., 2017, 2018). Translated to a monetary value, this threshold comes down to €12,400.

The second two-part model is based on the Pareto tail approach (as described in Paragraph 2.4.5). A log-rank, log-cost plot was modelled to determine the threshold. This plot is shown in Figure 7. The threshold was set at a log-cost value of 11, which marks the point at which the line becomes linear. This threshold value is equivalent to a monetary threshold of €59,874.

Figure 7: Plot of logarithm of rank against expenditure



Based on the costs for the year 2023, 17,736,376 observations and 10,000 bins. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

The third two-part model identifies high-cost users as individuals with healthcare costs equal to or above €100,000. This threshold is based on the absolute monetary threshold approach (which was discussed in Paragraph 2.4.1) and aligns with the research by Hamstra et al. (2023). The second criterion for high costs in the research by Hamstra (a minimum undercompensation of €50,000) will not be applied to identify high-cost users. This is because undercompensation already implies a fitted model, making this criterion suitable only for ex-post analysis rather than the ex-ante approach used in this study.

The percentiles method (which was discussed in Paragraph 2.4.2) will not be explored as a threshold value. This is because this method is sensitive to outliers. Extremely high costs, which can be seen as outliers, may disproportionately influence the threshold. This could result in a threshold that is not

representative for the group of high-cost users. Furthermore, the Lorenz curve method (which was discussed in Paragraph 2.4.4) will not be explored as a threshold value. By applying the top-percentage approach, a conceptually similar method to the Lorenz curve method will already be applied, as both approaches are based on percentages.

Perfect multicollinearity

When estimating the p-regression of the two-part models, there was a risk of perfect multicollinearity (as discussed in Paragraph 3.2.3). The top-percentage model was affected by this issue. When modelling the p-regression of this model, the ‘PCG50’ variable showed perfect multicollinearity. Therefore, this variable was omitted from the p-regression in this model to ensure that all observations were used in the prediction of the probability to incur high healthcare costs.

Cost transformation

To be able to compare the predicted healthcare costs across the two-part models, the LS regression model and the observed costs, the two-part model predictions were transformed. This was done by multiplying the individual predicted healthcare costs ($\hat{Y}_{3i}$) with the transformation factor. Table 4 shows the pre-transformation mean healthcare costs and the transformation factors applied to the two-part models. Furthermore, Table 4 shows the summarised results from all two-part models, the LS model and the 2026 model. This table shows that the mean healthcare costs are the same for each of these models, as a result of the cost transformation.

Table 4: Mean and standard deviation of actual healthcare spending and predicted healthcare spending according to different models

	Pre-transformation		Post-transformation	
	Mean	Transformation factor	Mean	Std. dev.
Actual healthcare costs	€2,834	N/A	€2,834	10,077
Predicted costs, LS model	€2,834	N/A	€2,834	5,643
Predicted costs, top-percentage model	€2,347	1.21	€2,834	6,087
Predicted costs, Pareto tail model	€2,693	1.05	€2,834	5,498
Predicted costs, absolute threshold model	€2,759	1.03	€2,834	5,457

Based on the cost for the year 2023, and 17,736,376 observations. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

4.3.2 Characteristics of high-cost users

Table 5 shows the size and age of the high-cost groups according to each model. As expected, the number of high-cost users decreases when the threshold increases. This results in the top-percentage model having a relatively large high-cost user group and the absolute threshold model having a relatively small high-cost user group. Within all two-part models, the high-cost user group includes a large proportion of individuals aged 75 years and older. In the top-percentage model, only 5.1% of all high-cost users are individuals between zero and four years old. As the threshold rises, the percentage of high-cost users between the ages of zero and four years old also rises. In the absolute threshold model, 14.1% of all high-cost users are between the ages of zero and four years old. This observation is also (less extremely) visible for high-cost users between the ages of five and seventeen years old.

Table 5: Percentage of high-cost individuals by age group

	Top-percentage model ($T \approx \text{€}12,400$)	Pareto tail model ($T \approx \text{€}59,874$)	Absolute threshold model ($T \approx \text{€}100,000$)
<i>Nr of high-cost users (HCU)</i>	886,818	104,133	41,380
<i>Percentage of HCU in the population</i>	5.0%	0.6%	0.2%
<i>% of HCU with age 0-4</i>	5.1%	9.1%	14.1%
<i>% of HCU with age 5-17</i>	3.0%	2.7%	3.9%
<i>% of HCU with age 18-39</i>	8.3%	5.2%	6.5%
<i>% of HCU with age 40-59</i>	16.7%	15.3%	15.3%
<i>% of HCU with age 60-74</i>	30.0%	31.8%	29.2%
<i>% of HCU with age 75+</i>	37.0%	36.0%	31.0%

Based on the cost for the year 2023. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

Table 6 shows the share of high-cost users who have specific risk characteristics. This table shows that as the threshold rises, the percentage of high-cost users who are in the top three classes for the MHC risk characteristic also rises. This implies that lower thresholds capture many individuals in the high-cost group who are not in one of the top three MHC classes. Furthermore, Table 6 shows that it is common for high-cost users to have a DCG and/or a PCG risk characteristic. A minimum of 71.5% of all high-cost users among all models has a PCG characteristic, implying that individuals who have high healthcare costs often use specific pharmaceuticals. Lastly, Table 6 shows that the application of a higher threshold results in a larger share of high-cost users who are male.

Table 6: Percentage of high-cost individuals by risk characteristic

	Top-percentage model ($T \approx \text{€}12,400$)	Pareto tail model ($T \approx \text{€}59,874$)	Absolute threshold model ($T \approx \text{€}100,000$)
<i>Nr of high-cost users (HCU)</i>	886,818	104,133	41,380
<i>% of HCU in top 3 risk classes for MHC</i>	4.8%	14.3%	17.3%
<i>% of HCU in top 3 risk classes for MCC</i>	6.1%	8.5%	8.0%
<i>% of HCU with at least one DCG</i>	55.1%	65.2%	62.9%
<i>% of HCU with at least one PCG</i>	73.8%	76.0%	71.5%
<i>% of male HCU</i>	47.3%	54.4%	56.4%
<i>% of female HCU</i>	52.7%	45.6%	43.6%

Based on the cost for the year 2023. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

Table 7 shows the actual healthcare costs of high-cost users in 2023. As expected, the mean healthcare costs of high-cost users are the highest in the absolute threshold model, as the threshold is the highest. In this model, high-cost users have mean healthcare costs of approximately €165,000. The interquartile range (IQR) and the median costs of each model rise as the threshold becomes higher. When comparing the mean costs to the median, it is worth noting that for all models, the mean costs are higher than the median. This indicates that the costs have a positively skewed distribution.

Table 7: Healthcare costs of high-cost individuals

	Top-percentage model ($T \approx \text{€}12,400$)	Pareto tail model ($T \approx \text{€}59,874$)	Absolute threshold model ($T \approx \text{€}100,000$)
<i>Nr. of high-cost individuals</i>	886,818	104,133	41,380
Mean costs	€31,154	€104,869	€165,360
Interquartile range	€18,552	€42,390	€61,771
Median costs	€20,986	€83,838	€132,609

Based on the cost for the year 2023. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

4.4 Evaluation and additional analyses

4.4.1 Evaluation of the estimation models

In Table 8, the three two-part models are evaluated based on the criteria from the WOR-assessment framework. This evaluation is done in comparison to the LS regression model. A summary of the evaluation is provided below, followed by a more detailed elaboration.

In terms of payment fit, the LS model and the absolute threshold model have the best results. In the LS model, two non-high-cost subgroups have the best payment fit of all models. In the absolute threshold model, two high-cost subgroups have the best payment fit of all models. However, the differences in payment fit between these models are relatively small. The top-percentage model shows a relatively big improvement in payment fit for the subgroup of enrollees who belonged to the top 15% of the highest costs.

The efficiency incentives of the two-part models are, given the fact that the same risk classes are used, similar to those of the LS regression model. Additional research is needed to assess how equalisation payments are allocated across morbidity criteria in the two-part models, as this may provide further insight into the efficiency incentives.

In terms of validity and measurability, the two-part models perform comparably to the LS regression model. This is because all models use the same risk adjusters. It remains unclear whether the results from the two-part models are as stable as the LS regression model, because the analyses were not repeated using data from different equalisation years.

Two-part models are more complex than the LS regression model. This is mostly due to the new payment weight calculation, which results in multiple payment weights per risk adjuster.

Table 8: The cohesion of the criterion, compared to the LS regression model

	Top-percentage model ($T \approx \text{€}12,400$)	Pareto tail model ($T \approx \text{€}59,874$)	Absolute threshold model ($T \approx \text{€}100,000$)
Payment fit	+/-	-	+/-
Efficiency	0	0	0
Validity, measurability & stability	0	0	0
Manageable complexity	-	-	-

+/- = The model has both advantages and disadvantages relative to the LS regression model, resulting in an overall neutral assessment.

0 = The model performs similarly to the LS regression model.

- = The model performs worse than the LS regression model.

Payment fit

The average result per subgroup, per model, is shown in Table 9. This table shows that the LS regression model performs best in terms of the payment fit for three out of the seven subgroups. Two of these subgroups consist of enrollees who are expected to have non-high healthcare costs (max deductible, 10% lowest premiums). However, for the subgroup with restrictive conditions, the payment fit is best in the absolute threshold model, even though this group is also expected to have non-high healthcare costs. The absolute threshold model performs best in two of the four subgroups in which high-cost users are expected (chronic condition, pharma costs). However, the subgroup of individuals with the 10% highest premiums is best compensated in the LS model. The top-percentage model performs best for the subgroup of enrollees who belonged to the top 15% of costs, relative to the LS model, with an improvement of €83. An explanation for this rather big improvement could be the overlap between the threshold-setting method and the subgroup. The high-cost user group consists of the 5% of individuals with the highest costs, and this subgroup consists of the 15% of individuals with the highest costs.

In terms of incentives for risk selection, Table 9 shows that the absolute threshold model and the LS model give the least incentives for risk selection, but for differing subgroups. The LS model provides the least incentives for risk selection among non-high cost subgroups. The absolute threshold model generally provides the weakest incentives for risk selection among high-cost subgroups. However, the LS model also improves one high-cost subgroup and vice versa. When comparing the absolute threshold model to the LS regression model, it is noticeable that the differences between the payment fit for each subgroup are relatively small.

When comparing the two-part models, the absolute threshold model performs best in terms of payment fit for six out of the seven evaluated subgroups. This indicates that the absolute threshold model gives the least incentives for risk selection among all two-part models.

Table 9: WOR-assessment framework, payment fit subgroup analysis

		LS regression model	Top-percentage model ($T \approx €12,400$)	Pareto tail model ($T \approx €59,874$)	Absolute threshold model ($T \approx €100,000$)
Subgroup	Frequency	Mean Financial Result	Mean Financial Result	Mean Financial Result	Mean Financial Result
<i>Presence of at least one chronic condition</i>	35.9%	-110	-157	-107	-95
<i>Enrollees who have had pharmacy costs for multiple years</i>	30.5%	-167	-210	-167	-152
<i>Enrollees who have chosen the maximum voluntary deductible</i>	8.1%	217	267	232	226
<i>Enrollees who have chosen an insurance contract with restrictive conditions</i>	24.4%	62	92	66	60
<i>Enrollees who have an insurance contract which falls under the 10% highest premiums</i>	4.6%	-109	-128	-123	-116
<i>Enrollees who have an insurance contract which falls under the 10% lowest premiums</i>	8.6%	102	134	109	103
Additional subgroups					
<i>Enrollees who, in 2020, belonged to the 15% of people with the highest somatic healthcare costs</i>	12.9%	-89	-6	-147	-109

Based on the costs for the year 2023. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

To further assess the payment fit of each model, Table 10 presents the statistical measures per model. The LS model has an R-squared value of 31.35, indicating that 31% of the variation in healthcare costs is explained by the model. None of the two-part models show an increase in the R-squared value. However, the differences between the R-squared values are relatively small. The absolute threshold model has an R-squared value of 31.22, showing only a 0.13 decrease relative to the LS model. When comparing the CPM values, it can be concluded that all models score comparably. Furthermore, the MAE is lowest in the top-percentage model, while the other models differ by a maximum of €5.

Table 10: Statistical measures

	LS regression model	Top-percentage model ($T \approx €12,400$)	Pareto tail model ($T \approx €59,874$)	Absolute threshold model ($T \approx €100,000$)
<i>Nr. of high-cost users</i>	<i>n.a.</i>	886,818	104,133	41,380
R-squared	31.35	30.05	31.04	31.22
CPM	35.33	35.40	35.26	35.31
Mean Absolute Error	€2,246	€2,243	€2,248	€2,246

Based on the cost for the year 2023. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

Efficiency

The efficiency incentives of the two-part models are comparable to those of the LS regression model. This is because the two-part models and the LS model use precisely the same risk adjusters. However,

to get a more comprehensive insight into the incentives from the two-part models, further research into the resulting payment weights is needed. This research should focus on the relationship between healthcare costs/healthcare use and the equalisation payments. This can, for example, be evaluated using power measures, as outlined by Douven et al. (2026).

In comparison to the 2026 risk equalisation model, a small shift in the efficiency incentives is expected. This is because the 2026 model uses ex-post compensation for high healthcare costs, and the two-part models only use ex-ante compensation. This ex-ante compensation makes insurers more risk-bearing, which enhances the efficiency incentives for insurers. However, since the ex-post compensation is only applied to a small group of individuals, this efficiency improvement has limited effects.

Validity, measurability & stability

The validity and measurability of the two-part models are comparable to those of the LS regression model. This is, as mentioned in the efficiency paragraph, because these models use the same risk adjusters. It remains uncertain whether the results from the two-part models are as stable as the LS regression model. This is because this research did not repeat the analyses on data from a different equalisation year. This could be done to assess the stability of two-part models.

Manageable complexity

Table 11 shows the complex elements of the two-part models. A two-part model enhances the complexity of the interpretation of results and the determination of contributions. The complexity of the result interpretation mostly stems from separate payment weights. In the LS regression model, each payment weight can be interpreted as the contribution or deduction of healthcare costs. This interpretation no longer holds in the two-part model. The payment weights for high costs in the two-part model are conditional on the probability of incurring high healthcare costs. This makes the model less intuitive to interpret. The enhanced complexity of the contribution determination stems from the new process to determine payment weights. Furthermore, because of the cost transformation, the two-part model becomes more complex, as this is a deviation from the regular process of contribution determination. The enhanced complexity negatively impacts the transparency and explainability of the model. Within the policy context of the Dutch risk equalisation system, this may create challenges for implementing the two-part model, as the interpretation of the payment weights is more difficult to communicate to stakeholders. Furthermore, the separate payment weights create an additional disadvantage when incorporating restrictions such as constrained regression into the model. Due to the payment weights, it remains unclear whether and how restrictions can be implemented.

Table 11: Manageable complexity

Process step	Expected effect on complexity, compared to the 2026 RE model	Explanation
The collection of source data	No difference	The same source data is used as the 2026 risk equalisation model
Preparation of execution files	No difference	The same execution files are used as the 2026 risk equalisation model
Characteristic derivation	No difference	No new risk characteristic is added to the model
Interpretation of results	Enhances complexity	Separate payment weight for 'high costs', which is hard to interpret
Application of ex-post measures	No difference	The LS regression model and two-part models both do not use ex-post measures
Calculation of payment weights	No difference	No small risk classes or unstable risk class populations were added to the model
Contribution determination	Enhances complexity	Deviation from the regular process for the determination of payment weights and the use of cost transformation

4.4.2 Stepwise regression analysis

A stepwise regression analysis was performed to gain insight into which variables have a significant impact on the two-part models. This additional analysis was performed using the absolute threshold model. This model has been selected based on the results shown in Tables 9 and 10. Table 9 showed that for six out of the seven evaluated subgroups, the absolute threshold model had the best payment fit of all two-part models. The absolute threshold model also resulted in the least loss of payment fit for the non-high-cost subgroups. Furthermore, all two-part models were comparable in terms of the statistical measures shown in Table 10.

Table 12 shows the non-significant variables at the 1% significance level for each regression in the absolute threshold model. For each regression, the R-squared values relating to the stepwise regression and the non-stepwise regression are comparable. This shows that the model keeps its predictive performance when omitting several variables. It is worth noting that the R-squared values differ across the regressions. The Y_1 regression has the highest R-squared value of all regressions, indicating that healthcare costs below the threshold can be estimated most accurately. Furthermore, the R-squared value of the p-regression suggests that the probability of incurring high healthcare costs can also be estimated quite well. However, the Y_2 regression has a very low R-squared value, indicating that the height of high healthcare costs is difficult to predict.

In the Y_1 regression, the fewest risk adjuster classes are omitted. This is expected because this regression mostly resembles the currently used methodology of the risk equalisation model. There is no clear pattern in the risk adjuster classes which are omitted, because only a limited number of classes are omitted, which are spread across eight different risk adjusters.

Most risk adjuster classes are omitted in the Y_2 -regression, which suggests that not all risk adjuster classes are useful for predicting the height of high healthcare costs. This is consistent with the low R-squared value of the regression. The PCG risk adjuster classes make up for 27% of the omitted risk adjuster classes in this regression. Half of all PCG risk adjuster classes were omitted in this regression.

The omitted classes are evenly distributed across all PCG classes, with only the top four classes not being omitted. This suggests that the use of specific pharmaceuticals outside these four top classes does not significantly influence the height of high healthcare costs. Furthermore, 26% of all omitted risk adjuster classes stem from the age/sex variable. These omitted classes include men and women who are 18 years and older. This suggests that being aged 18 years or above does not play a significant role in determining the height of high healthcare costs. Lastly, risk adjuster classes based on the source of income make up for 22% of all omitted classes. The omitted classes include social assistance recipients, individuals receiving disability benefits and self-employed individuals. This suggests that obtaining income through one of these sources does not have a significant impact on the height of high healthcare costs.

The p-regression omits PCG risk adjuster classes, which account for 32% of all omitted classes. The omitted classes are evenly distributed among the PCG classes, with only the top four classes not being omitted. This suggests that the use of a specific pharmaceutical, outside these four classes, does not significantly influence the probability of incurring high healthcare costs. Furthermore, 23% of all omitted classes in this regression stem from the income-based risk adjuster. The omitted risk classes include students, highly educated individuals and self-employed individuals. This suggests that obtaining income from one of these sources does not have a large impact on the probability of incurring high healthcare costs. A list of all omitted risk classes from the stepwise regressions is shown in Appendix 4.

Table 12: Non-significant variables at the 1% level in the absolute threshold model

Regression	R-squared value stepwise	R-squared value non-stepwise	Number of omitted risk adjuster classes	Omitted variables (amount of risk classes x risk characteristic)
$\hat{Y}_{1i}$ (Equation 15)	40.65	40.65	13	3 x SOI 3 x SES 2 x Region 1 x PPA 1 x Age/Sex 1 x HCA 1 x MHC 1 x PCG
$\hat{Y}_{2i}$ (Equation 16)	8.92	8.92	92	25 x PCG 24 x Age/Sex 20 x SOI 8 x Region 3 x IP 3 x SES 3 x MCH 2 x PPA 1 x HCA 1 x MHC 1 x DCG 1 x MCHC
$\hat{p}_i$ (Equation 17)	35.35	35.38	77	25 x PCG 4 x SES 18 x SOI 2 x IP 12 x Age/Sex 8 x Region 3 x DCG 2 x PPA 1 x HCA 1 x MHC 1 x PTD

Based on the cost for the year 2023, and 17,736,376 observations. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

4.4.3 The two-part model compared to the 2026 risk equalisation model

Table 13 shows the Mean Financial Result per subgroup for the best-performing two-part model, the 2026 model with and without constrained regression (CR) and the LS model. The table shows that the differences between the 2026 model without constrained regression and the two-part model are relatively limited. However, larger differences remain between the two-part model and the 2026 model with constrained regression. This indicates that a two-part model cannot perform as a substitute for constrained regression in the equalisation model. To further align the results from the absolute threshold model with the 2026 equalisation model with CR, CR can be applied to the Y_1 regression of the two-part model. This is because the Y_1 regression most resembles the 2026 risk equalisation model. The comparison between the two-part model and the 2026 equalisation model without CR does indicate that a two-part model, in broad terms, could serve as an alternative modelling approach. This is because there are relatively small differences between the outcomes.

Table 13: Payment fit, subgroup analysis 2026 risk equalisation model

		LS regression model	Absolute threshold model ($T \approx €100,000$)	2026 model without constrained regression ²	2026 model with constrained regression ¹
Subgroup	Frequency	Mean Financial Result	Mean Financial Result	Mean Financial Result	Mean Financial Result
<i>Presence of at least one chronic condition</i>	35.9%	-110	-95	-87	28
<i>Enrollees who have had pharmacy costs for multiple years</i>	30.5%	-167	-152	-151	0 *
<i>Enrollees who have chosen the maximum voluntary deductible</i>	8.1%	217	226	208	129
<i>Enrollees who have chosen an insurance contract with restrictive conditions</i>	24.4%	62	60	61	23
<i>Enrollees who have an insurance contract which falls under the 10% highest premiums</i>	4.6%	-109	-116	-108	-86
<i>Enrollees who have an insurance contract which falls under the 10% lowest premiums</i>	8.6%	102	103	102	58

Based on the cost for the year 2023. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

¹ The 2026 risk equalisation model, without the ex-post high-cost compensation. Source: Gupta Strategists (2025b)

² The 2026 risk equalisation model, without ex-post high-cost compensation and constrained regression. Source: Gupta Strategists (2025b)

* Constrained regression is applied to the variable which indicates this subgroup

To further compare the two-part model to the 2026 model, Table 14 shows the statistical measures. There are only relatively small differences between the statistical measures from the 2026 models and the two-part model. This indicates that a two-part model has a statistical performance that is broadly comparable to that of the 2026 model.

Table 14: Statistical measures

	LS regression model	Absolute threshold model ($T \approx €100,000$)	2026 model without constrained regression ²	2026 model with constrained regression ¹
R-squared	31.35	31.22	31.32	31.30
CPM	35.33	35.31	35.29	35.29

Based on the cost for the year 2023. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.

¹ The 2026 risk equalisation model, without the ex-post high-cost compensation. Source: Gupta Strategists (2025b)

² The 2026 risk equalisation model, without ex-post high-cost compensation and constrained regression. Source: Gupta Strategists (2025b)

5. Discussion

In this chapter, the findings of this study are discussed. First, a summary of the key findings is provided, in which the sub-questions and the study's contribution to the literature are discussed. Next, strengths, limitations and recommendations for future research are presented. Third, the policy implications of the results found in this study are discussed. Lastly, a conclusion of the study is provided by answering the research question of the thesis.

5.1 Summary of key findings

The first sub-question addressed in this study is: *How can high healthcare costs be defined and what are the key risk characteristics of individuals with high healthcare costs?* High healthcare costs can be defined by using a cost threshold. Three thresholds were assessed during this research: the top-percentage threshold, the Pareto tail threshold and the absolute monetary threshold. Each of these thresholds stems from a different threshold-setting method. Individuals with high healthcare costs, defined as equal to or above €100,000, have specific key risk characteristics. Individuals with high healthcare costs are often 75 years or older, and between the ages of 0 and 4. When comparing all thresholds, it is noticeable that the share of 0-to 4-year-olds rises when the high-cost threshold rises. In addition to age, a key characteristic of high-cost users is the likelihood of having at least one Pharmacy Cost Group (PCG) and/or Diagnostic Cost Group (DCG) risk adjuster. This shows that high-cost users often use specific pharmaceuticals and/or have had specific hospital treatments in the prior year. Lastly, it is noticeable that high-cost users are more likely to be in the top three risk classes for the Multiple Year High Cost risk adjuster, as the threshold increases.

The second sub-question that is examined in this study is: *To what extent does a two-part model improve the outcome of the standard evaluation measures relative to the least squares method for somatic care?* Multiple standard evaluation measures were applied during the data analyses. In terms of the payment fit, it can be concluded that two-part models improve the payment fit for specific subgroups. It depends on the threshold, which subgroups have an improved payment fit relative to the LS regression model. The absolute threshold model improves the payment fit for individuals who have had pharmacy costs for multiple years, chronically ill individuals and individuals who have chosen an insurance contract with restrictive conditions. The Pareto tail model also improves the payment fit for individuals who are chronically ill. The top-percentage model improves the payment fit for individuals who, in 2020, belonged to the top 15% highest cost individuals. Overall, it can be concluded that the absolute threshold model results in the best payment fit among the two-part models, as this model improves most subgroups. A possible explanation for this lies in the threshold used in each two-part model. The absolute threshold model uses the highest threshold value, resulting in a smaller high-cost group. Therefore, the high-cost user group is likely to be less heterogeneous, which could enhance the payment fit. When comparing the absolute threshold model with the least squares method, it can be concluded that the absolute threshold model improves the payment fit for individuals who are chronically ill, individuals who have an insurance contract with restrictive conditions and individuals who have had pharmacy costs for multiple years. However, at the same time, the payment fit decreases in other subgroups. The outcomes of the statistical measures from the LS regression model are comparable to those of the two-part models. Therefore, it can be concluded that the statistical measures are not improved when using a two-part model. When comparing the efficiency incentives of the LS model to the two-part models, it can be concluded that a two-part model does not improve nor decrease the efficiency incentives of the model. Furthermore, a two-part model does not improve or decrease the validity and measurability of the model. However, a two-part

model does have an increased complexity relative to the LS regression model. This is because of the separate payment weights for high healthcare costs, which are dependent on the probability of incurring high healthcare costs. It remains uncertain whether the results from the two-part models are as stable as the LS-regression model.

The third sub-question that is examined in this study is: *To what extent does a two-part model reduce under- and overcompensation for high-cost individuals and chronic disease subgroups relative to the least squares method for somatic care?* During the analyses, four high-cost subgroups were evaluated: individuals who are chronically ill, individuals who have had pharmacy costs for multiple years, individuals with the 10% highest premiums and individuals who belonged to the top 15% highest costs. In the LS model, these subgroups are undercompensated. The top-percentage model decreases the undercompensation for the subgroup with the 15% highest costs by €83. The absolute threshold model decreases the undercompensation for the subgroup with individuals who are chronically ill and for the subgroup with pharmacy costs for multiple years. The Pareto tail model also decreases the undercompensation for individuals who are chronically ill. This shows that all two-part models reduce the undercompensation for high-cost individuals, although the specific subgroups and magnitude of the reduction differ. Furthermore, in the LS regression model, the subgroup of individuals with a chronic disease is undercompensated. The absolute threshold model reduces this undercompensation by €15. The Pareto tail model reduces this undercompensation by €3. This shows that the absolute threshold model and the Pareto tail model reduce the undercompensation for the subgroup of individuals with a chronic disease.

The fourth sub-question examined in this study is: *What are the pros and cons of two-part models and potential alternatives for predicting high healthcare costs?* Two-part models have the advantage of increasing the payment fit for various subgroups. However, a disadvantage of the two-part model is that, in turn, other subgroups have a reduced payment fit. Furthermore, an advantage of the two-part models is that, in comparison to the 2026 risk equalisation model, the two-part model slightly improves the efficiency incentives of the model. This is because the ex-post compensation for high costs could be reduced due to better estimation of high healthcare costs. A disadvantage of the two-part model is the increased complexity because of the calculation of separate payment weights. Potential alternatives for predicting high healthcare costs, which have been discussed in Paragraph 2.3, include the application of a three-part model with an even higher threshold for extremely high costs. Furthermore, alternative threshold values could be applied to assess whether an even higher threshold further improves the predictive performance of the two-part model.

The results from this study add to the small amount of existing literature regarding this topic. In particular, before this research, no two-part model with a non-zero cost threshold had been researched in terms of risk equalisation. Therefore, the results of this research provide new insights into the application of two-part models. This research builds on prior work on high-cost thresholds and demonstrates how these thresholds can be applied in two-part equalisation models (Karlsson et al., 2024; Hamstra et al., 2023; Wammes et al., 2017, 2018). The results show that the absolute monetary threshold, originating from Hamstra et al. (2023), results in the best payment fit for multiple (high-cost) subgroups. Furthermore, the results from this research give a more specific insight into the characteristics of individuals with high healthcare costs. Research by Hamstra et al. (2023) and Wammes et al. (2017, 2018) demonstrated characteristics of high-cost users. This research specifies these characteristics by further defining high-cost users according to risk adjustment characteristics from the Dutch 2026 risk equalisation model. This research has shown that high-cost users often have a Pharmacy Cost Group and/or Diagnostic Cost Group risk adjuster.

5.2 Strengths, limitations & directions for future research

When interpreting the results from this research, the strengths and limitations of this research should be taken into account. This paragraph discusses the most important strengths and limitations, supplemented with recommendations for future research.

The use of real-world data

A strength of this study is the use of real-world data from 17 million Dutch citizens with basic health insurance in 2023. This data is also used in the Dutch 2026 risk equalisation model, which makes the results from this study applicable to the whole population. The data used only entails data on costs from 2023. Therefore, no stability tests were done with data from other cost years. In future research, the data analyses could be repeated with data on other cost years. This makes it possible to further assess the stability of the results from the estimated models. This aligns with the WOR-assessment framework, as this gives more insight into the validity, measurability and stability criteria of the model (Ministerie van Volksgezondheid, Welzijn en Sport, 2024).

Omission of restrictions

Paragraph 3.4 of the thesis discussed the omission of several restrictions from the 2026 risk equalisation model. The omission of these restrictions improved the comparability between the simulated LS regression model and the two-part models. As a result, the differences in outcomes between the models could be attributed more directly to the modelling approach rather than the restrictions. However, this choice reduced the representativeness of the models with respect to the 2026 risk equalisation model. Furthermore, removing these restrictions resulted in non-intuitive payment weights, making these payment weights unsuitable for direct interpretation. This led to a limited comparability between the payment weights of the estimated models and the 2026 model. Because of this, the efficiency incentives on cost-containment could not be properly assessed. However, since this research focused on the prediction of healthcare costs instead of payment weights, these consequences had a limited effect on the results. This ensured that the predicted healthcare costs were sufficiently comparable to the 2026 equalisation model without constrained regression, enabling meaningful conclusions. This was illustrated in Paragraph 4.4.3, which showed that the results from the LS regression, the 2026 model without CR, and the absolute threshold model were generally comparable.

Future research could assess the implementation of restrictions into the two-part models. However, it remains unclear whether this is feasible and, if so, how the restrictions should be incorporated within a two-part model. This additional analysis could give more insight into the efficiency incentives of the two-part models and enhance alignment with the methodology of the WOR-assessment framework.

Cost transformation in the two-part models

Paragraph 3.2.3 of the thesis discussed the transformation of healthcare costs to align the mean healthcare costs of the two-part models with the mean healthcare costs of the 2026 model and the mean observed costs. This transformation improved the comparability between the models. However, it is uncertain whether the transformation actually improved the prediction accuracy. This is because the transformation method assumed the prediction errors to be constant across the distribution. This may not hold as the distribution is skewed. Nevertheless, the transformation is considered appropriate in this study as the goal was to compare model performance, and the transformation made the models comparable. However, it is important to note that this research did not focus on the implications of

cost transformation for research on and the execution of risk equalisation. This can be a point of attention in possible future research.

Threshold setting

The empirical analysis was conducted using three different thresholds for having high healthcare costs. The results showed that the use of the highest threshold (€100,000) resulted in a two-part model which had the best payment fit for most subgroups. This raises the question of whether an even higher threshold results in a better payment fit for more subgroups. The theory on threshold setting showed more methods which can be used to set a high-cost threshold, such as the Lorenz curve method. This method could be used to set a different threshold. However, other methods could also be applied, as long as the threshold is higher than €100,000. Furthermore, Paragraph 2.3.4 discussed an expansion of the two-part model by Karlsson et al. (2024). This gives the opportunity to expand the two-part model to a three-part model in which an additional threshold is used to capture individuals with extremely high costs. It should be noted that this extension differs from the Karlsson et al. method, as the current two-part model already includes a high-cost threshold. Therefore, it is unclear whether this extension will add predictive power. Moreover, the expansion of the two-part model involves a trade-off, since the group with extremely high costs will be relatively small, as this is a group of the current 0.2% of the population which is a high-cost user.

Predicting (the probability of incurring) high healthcare costs

The stepwise regression analysis showed that the risk adjuster classes used in the p - and Y_2 regressions are by no means all significant. This is further supported by the relatively low R-squared value of the Y_2 regression. However, this is rather logical, as the risk adjusters used in the p - and Y_2 regressions are not intended to estimate the probability of incurring high healthcare costs or the height of high healthcare costs. This shows a possible mismatch between the theory and the model. Therefore, an implication for future research is to explore new data sources (and therefore risk adjusters) which are available to incorporate into these two regressions. This way, the model could better predict the probability of incurring high healthcare costs and the height of high healthcare costs. This possible improvement of the two-part model could lead to an improvement in the estimated healthcare costs.

Evaluation criteria

This research focused on a limited set of evaluation measures. Naturally, many other measures and subgroups are possible for the evaluation of the models. In future research, more subgroups could be assessed in terms of payment fit. This includes subgroups based on specific conditions. This can be done by applying the Nivel dataset, which provides data on medical conditions of approximately 1.5 million individuals in the Netherlands in 2022. Furthermore, subgroups based on self-reported health could be assessed in terms of payment fit. This can be done by using the GGD health monitor dataset. An example of the application of these subgroups is illustrated in van Kleef et al. (2026).

Efficiency incentives

Due to feasibility constraints, no power measures were calculated to assess the efficiency incentives of the two-part models. Future research could apply power measures to examine the relationship between healthcare costs/healthcare use and the equalisation payments. This additional analysis could provide more insight into how equalisation payments from the two-part models are distributed across morbidity criteria and whether the payment weights accurately reflect actual treatment costs in practice.

Alternative estimation methods

This thesis focused on two-part models to estimate healthcare costs. The sensitivity of the results to the threshold could further be explored outside a two-part model. A two-part model needs a threshold to determine high healthcare costs; other estimation methods do not have this obstacle. Quantile regression can be used to estimate high healthcare costs without needing a high-cost threshold. According to Agung (2021), quantile regression is the extension of linear regression and can be used when the conditions of a linear regression are not applicable, as is the case with skewed healthcare costs. Instead of defining a high-cost threshold, quantile regression models the upper percentiles of the cost distribution. Lorenz (2017) studied the use of quantile regression in risk adjustment and concluded that quantile regression-based risk adjustment performs well in highly skewed distributions.

The limited predictive performance of the Y_2 regression may not only result from the limited risk adjusters, but also from the choice of the estimation model. As mentioned in Paragraph 2.3.1, LS regressions are bound to the normality assumption, which does not hold in a skewed distribution. Generalised Linear Models (GLM) are an alternative to the LS regression, as they better accommodate heavy right tails (Blough et al., 1999). A study by Blough et al. (1999) researched the use of GLM in the second part of a risk equalisation two-part model. This two-part model was based on a cost/no-cost threshold. The study concluded that the application of a GLM better fits the skewed cost distribution and results in the statistical properties being better reflected. Applying a GLM instead of an LS regression in the Y_2 regression could therefore lead to an improvement in the estimation of high healthcare costs.

5.3 Policy implications

The findings of this study show that a two-part model improves the payment fit for different subgroups, among which are high-cost users. However, the results also show that the improvement in payment fit for some subgroups comes with a decrease in payment fit for other subgroups. In the risk equalisation policy context, this means that the use of a two-part model shifts the selection incentives on a subgroup level. The use of a two-part model results in fewer selection incentives for some subgroups, but in more selection incentives for other subgroups. This means that overall, the use of a two-part model involves a clear trade-off. Furthermore, the two-part model is more complex than an LS regression model because it requires separate payment weight estimations. As a result, there is also a trade-off between the increased complexity of the model and the limited improvement in payment fit for certain subgroups. These trade-offs should be discussed by the relevant stakeholders, such as health insurers and the Ministry of Health, Welfare and Sport.

In addition, the results from this research show that the 2026 risk equalisation model and the LS regression model work quite well to predict healthcare costs. A more sophisticated equalisation model, such as the two-part approach, only improves the payment fit for some subgroups while, in turn, giving a worse payment fit for other subgroups. This result suggests that there is only a limited potential for alternative estimation models in risk equalisation.

Furthermore, this study shows that the application of a two-part model does not result in a substantive improvement in the estimation of healthcare costs for individuals with a predictable profit or loss. A positive implication of this limitation is that it prevents insurers from identifying these subgroups by applying the model themselves. As a result, two-part models do not enable insurers to strategically target these groups.

5.4 Conclusion

This section provides an answer to the research question of this thesis: *How do two-part models perform relative to the least squares method in predicting high healthcare costs for somatic care?* The performance of a two-part model is dependent on the threshold which is applied to determine high healthcare costs. When applying a fitting threshold, a two-part model improves the payment fit for individuals with high healthcare costs for somatic care, relative to the least squares method. The results from the analyses indicate that the absolute monetary threshold results in the best payment fit for high-cost users. When applying the absolute threshold model, two out of the four evaluated high-cost subgroups show an improved payment fit relative to the LS regression method. This entails the subgroup of individuals who are chronically ill and the subgroup of individuals who have had pharmacy costs for multiple years. The top-percentage model improves the payment fit for enrollees who belonged to the top 15% individuals with the highest costs, relative to the LS regression model. The Pareto tail model improves the payment fit for individuals who are chronically ill. This shows that the performance of a two-part model, relative to the LS regression model, is dependent on the set high-cost threshold. The two-part models improve the prediction of high healthcare costs. However, these models perform worse for other subgroups. It is important to note that the differences in payment fit for the evaluated subgroups are relatively small. Therefore, it can be concluded that two-part models with an appropriate threshold, in comparison to LS regression, improve the prediction of high healthcare costs for somatic care to a certain limit, but, at the same time, give a worse prediction for other subgroups.

Two-part models, in comparison to the LS regression model, are relatively complex. This means that the application of a two-part model will be more complex than the application of the least squares method. In terms of efficiency, validity and measurability, the two-part model has a comparable performance to the least squares method. Therefore, it can be concluded that two-part models do not improve nor worsen the efficiency, validity and measurability of the equalisation model, in comparison to the least squares method. However, two-part models do perform worse in terms of complexity, relative to the least squares method.

6. References

- Agung, I. G. N. (2021). *Quantile Regression* (1st ed). John Wiley & Sons, Ltd. <https://doi.org/10.1002/9781119714958>
- Armstrong, J. (2018). Health Plan Payment in Ireland. In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets* (pp. 331–364). Academic Press. <https://doi.org/10.1016/B978-0-12-811325-7.00012-9>
- Beck, K., Kauer, L., McGuire, T. G., & Schmid, C. P. R. (2020). Improving risk-equalization in Switzerland: Effects of alternative reform proposals on reallocating public subsidies for hospitals. *Health Policy*, 124(12), 1363–1367. <https://doi.org/10.1016/j.healthpol.2020.08.011>
- Belotti, F., Deb, P., Manning, W. G., & Norton, E. C. (2015). Twopm: Two-Part Models. *The Stata Journal*, 15(1), 3–20. <https://doi.org/10.1177/1536867X1501500102>
- Bes, R., Wendel, S., Jong, J. (2012). Het vertrouwensprobleem van zorgverzekeraars. *ESB: Economisch Statistische Berichten*, 97(4647), 676–677. <https://www.nivel.nl/nl/publicatie/het-vertrouwensprobleem-van-zorgverzekeraars>
- Bes, R., Curfs, E., Groenewegen, P., Jong, J. (2016). Vertrouwen in verzekeraar belangrijk voor acceptatie van selectief contracteren. *KIZ: Tijdschrift over Kwaliteit en Veiligheid in Zorg*, 4-7. <https://www.nivel.nl/nl/publicatie/vertrouwen-verzekeraar-belangrijk-voor-acceptatie-van-selectief-contracteren>
- Blough, D. K., Madden, C. W., & Hornbrook, M. C. (1999). Modeling risk using generalized linear models. *Journal of Health Economics*, 18(2), 153–171. [https://doi.org/10.1016/S0167-6296\(98\)00032-0](https://doi.org/10.1016/S0167-6296(98)00032-0)
- Buchner, F., Goepffarth, D., & Wasem, J. (2013). The new risk adjustment formula in Germany: Implementation and first experiences. *Health Policy*, 109(3), 253–262. <https://doi.org/10.1016/j.healthpol.2012.12.001>
- Chicco, D., Warrens, M. J., & Jurman, G. (2021). The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation. *PeerJ Computer Science*, 7, e623. <https://doi.org/10.7717/peerj-cs.623>
- Douven, R. (2000). *Regulated competition in health insurance markets*. The Netherlands Bureau for Economic Policy Analysis. <https://www.cpb.nl/en/publication/regulated-competition-health-insurance-markets>
- Douven, R. (2004). *Risk adjustment in the Netherlands*. The Netherlands Bureau for Economic Policy Analysis. <https://www.cpb.nl/system/files/cpbmedia/publicaties/download/risk-adjustment-netherlands-analysis-insurers-health-care-expenditures.pdf>
- Douven, R., Brammli, S., Buchner, F., Kauer, L., van Kleef, R., McGuire, T., Paolucci, F., Reitsma, M., Rose, S., Schmid, C.P.R., & Wasem, J. (2026). *Power of volume and price incentives in health plan payment models: A country comparison* (EsCHER Working Paper Series No. 2026003). Erasmus University Rotterdam. <https://www.eur.nl/en/research/research-groups-initiatives/escher/research/working-papers>
- Dziegielewska, C., Talarico, R., Imsirovic, H., Qureshi, D., Choudhri, Y., Tanuseputro, P., Thompson, L. H., & Kyeremanteng, K. (2021). Characteristics and resource utilization of high-cost users in the intensive care unit: A population-based cohort study. *BMC Health Services Research*, 21, 1–10. <https://doi.org/10.1186/s12913-021-07318-y>
- Eijkenaar, F., & Van Vliet, R. C. J. A. (2017). Improving risk equalization for individuals with persistently high costs: Experiences from the Netherlands. *Health Policy*, 121(11), 1169–1176. <https://doi.org/10.1016/j.healthpol.2017.09.007>
- Enthoven, A. C. (1993). The history and principles of managed competition. *Health Affairs*, 12, 24–48. https://doi.org/10.1377/hlthaff.12.suppl_1.24
- Espinosa, O., Buitrago, G., Bejarano, V., & Castaño, R. (2025). Lorenz curves of urban and rural health care expenditure in Colombia: Analysis of a cohort of 4.6 million patients. *Health Economics Review*, 15(1), 74. <https://doi.org/10.1186/s13561-025-00662-3>
- Field, M. J., & Jette, A. M. (2007). Risk Adjustment of Insurance Premiums in the United States and Implications for People with Disabilities. In *The Future of Disability in America*. National Academies Press. <https://www.ncbi.nlm.nih.gov/books/NBK11417/>

- Gupta Strategists. (2025a). *WOR 1277—Definitieve rapportage Normbedragen RV 2026*. Ministerie van Volksgezondheid, Welzijn en Sport. <https://www.zorginstituutnederland.nl/documenten/2025/09/30/wor-1277---definitieve-rapportage-normbedragen-rv-2026>
- Gupta Strategists. (2025b). *WOR 1275—Overall Toets, Onderzoek risicoverevening 2026*. Ministerie van Volksgezondheid, Welzijn en Sport.
- Hamstra, G., Borg, S., Mol, M., Boer, J., & Stam, P. (2023). *Compensatie extreem dure behandelingen, Eindrapportage WOR 1137*. Equalis. <https://open.overheid.nl/documenten/818b0201-2dae-462e-bc48-50892ff6a611/file>
- Henriquez, J., van Kleef, R. C., Matthews, A., McGuire, T., & Paolucci, F. (2023). *Combining Risk Adjustment with Risk Sharing in Health Plan Payment Systems: Private Health Insurance in Australia* (Working Paper No. 31052). National Bureau of Economic Research. <https://doi.org/10.3386/w31052>
- Huang, Y., Zheng, Y., Pan, X., & Zhou, L. (2025). Stepwise regression algorithm based on the ordered pair of normalized real numbers framework. *The Journal of Supercomputing*, 81(8), 900. <https://doi.org/10.1007/s11227-025-07369-6>
- Ismail, I., Stam, P. J. A., Portrait, F. R. M., van Witteloostuijn, A., & Koolman, X. (2024). Addressing unanticipated interactions in risk equalization: A machine learning approach to modeling medical expenditure risk. *Economic Modelling*, 130, 106564. <https://doi.org/10.1016/j.econmod.2023.106564>
- Jansen, M., Voorrips, L., & Peters, M. (2025). *Wie hebben de meeste zorgkosten?* Centraal Bureau voor de Statistiek. <https://www.cbs.nl/nl-nl/longread/statistische-trends/2025/wie-hebben-de-meeste-zorgkosten-/3-resultaten>
- Kapitula, L. R. (2015). *When two are better than one: Fitting two-part models using SAS®* (Paper 3600-2015). Grand Valley State University. <https://support.sas.com/resources/papers/proceedings15/3600-2015.pdf>
- Karlsson, M., Wang, Y., & Ziebarth, N. R. (2024). Getting the right tail right: Modeling tails of health expenditure distributions. *Journal of Health Economics*, 97, 102912. <https://doi.org/10.1016/j.jhealeco.2024.102912>
- Karunayawong, P., Gaewkhiew, P., Sarajan, M. H., Boonma, C., Butchon, R., Sukmanee, J., Saeraneesopon, T., Teerawattananon, Y., & Isaranuwachai, W. (2024). High-cost users still came to hospitals during the COVID-19 pandemic during first wave data in Thailand: Secondary data analysis. *BMC Public Health*, 24(1), 2917. <https://doi.org/10.1186/s12889-024-20325-y>
- Kauer, L., McGuire, T. G., & Beck, K. (2020). Extreme under and overcompensation in morbidity-based health plan payments: The case of Switzerland. *Health Policy*, 124(1), 61–68. <https://doi.org/10.1016/j.healthpol.2019.11.008>
- Leao, D. L. L., Pavlova, M., & Groot, W. (2023). Risk selection reduces efficiency of value-based healthcare. *The International Journal of Health Planning and Management*, 38(5), 1088–1096. <https://doi.org/10.1002/hpm.3648>
- Lorenz, N. (2017). Using Quantile and Asymmetric Least Squares Regression for Optimal Risk Adjustment. *Health Economics*, 26(6), 724–742. <https://doi.org/10.1002/hec.3352>
- Malehi, A. S., Pourmotahari, F., & Angali, K. A. (2015). Statistical models for the analysis of skewed healthcare cost data: A simulation study. *Health Economics Review*, 5(1), 11. <https://doi.org/10.1186/s13561-015-0045-7>
- Mauguen, A., & Begg, C. B. (2016). Using the Lorenz Curve to Characterize Risk Predictiveness and Etiologic Heterogeneity. *Epidemiology*, 27(4), 531–537. <https://doi.org/10.1097/EDE.0000000000000499>
- McGuire, T. G., Schillo, S., & van Kleef, R. C. (2021). Very high and low residual spenders in private health insurance markets: Germany, The Netherlands and the U.S. Marketplaces. *The European Journal of Health Economics*, 22(1), 35–50. <https://doi.org/10.1007/s10198-020-01227-3>
- Miller, A., Panneerselvam, J., & Liu, L. (2022). A review of regression and classification techniques for analysis of common and rare variants and gene-environmental factors. *Neurocomputing*, 489, 466–485. <https://doi.org/10.1016/j.neucom.2021.08.150>
- Minister van Volksgezondheid, Welzijn en Sport. (2023). *Herziening Zorgstelsel; vormgeving risicovereveningsmodel 2024*. <https://zoek.officielebekendmakingen.nl/kst-29689-1204.html>
- Ministerie van Volksgezondheid, Welzijn en Sport. (2017). *Beschrijving van het risicovereveningssysteem van de Zorgverzekeringswet*. <https://open.overheid.nl/documenten/ronl-a2fad2532fb0ec5984f8a9308f0ee22d10230fec/pdf>
- Ministerie van Volksgezondheid, Welzijn en Sport. (2024). *Toetsingskader 2024 en verder (WOR 1234)*.

- Ministerie van Volksgezondheid, Welzijn en Sport. (2025a). *Besluitvorming risicovereveningsmodel 2026*. <https://open.overheid.nl/documenten/077dd762-806e-49db-bd63-966ba09ef6fb/file>
- Ministerie van Volksgezondheid, Welzijn en Sport. (2025b). *Staatscourant regeling risicoverevening*. <https://zoek.officielebekendmakingen.nl/stcrt-2025-32634.html#d17e1443>
- Nederlandse Zorgautoriteit. (2015). *Verhouding tussen de aanvullende verzekering en de basisverzekering bij de sturing van verzekerden*. https://www.eerstekamer.nl/overig/20150630/verhouding_tussen_de_aanvullende
- Netherlands Authority for Consumers and Markets. (2016). *Competition in the Dutch Health Insurance Market*. https://www.acm.nl/sites/default/files/old_publication/publicaties/16129_competition-in-the-dutch-health-insurance-market.pdf
- 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>
- PwC. (2024). *Monitor Risicoverevening 2019*. <https://www.rijksoverheid.nl/documenten/rapporten/2024/04/30/monitor-risicoverevening-2019>
- Reich, N. G., Lessler, J., Sakrejda, K., Lauer, S. A., Iamsirithaworn, S., & Cummings, D. A. T. (2016). Case study in evaluating time series prediction models using the relative mean absolute error. *The American Statistician*, 70(3), 285–292. <https://doi.org/10.1080/00031305.2016.1148631>
- Reitsma, M. B., McGuire, T. G., & Rose, S. (2025). Algorithms to Improve Fairness in Medicare Risk Adjustment. *JAMA Health Forum*, 6(8), e252640. <https://doi.org/10.1001/jamahealthforum.2025.2640>
- Risk Adjustment Network. (n.d.). *Foundation*. <https://riskadjustment.net/#foundation>
- 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>
- Schmidt, A. F., & Finan, C. (2018). Linear regression and the normality assumption. *Journal of Clinical Epidemiology*, 98, 146–151. <https://doi.org/10.1016/j.jclinepi.2017.12.006>
- Schokkaert, E., & Van de Voorde, C. (2000). Risk adjustment and the fear of markets: The case of Belgium. *Health Care Management Science*, 3(2), 121–130. <https://doi.org/10.1023/A:1019089223462>
- Schut, E., & Varkevisser, M. (2016). *Economie van de gezondheidszorg*. Bohn Stafleu van Loghum. <https://doi.org/10.1007/978-90-368-1314-3>
- Shmueli, A., Chernichovsky, D., & Zmora, I. (2003). Risk adjustment and risk sharing: The Israeli experience. *Health Policy*, 65(1), 37–48. [https://doi.org/10.1016/S0168-8510\(02\)00115-X](https://doi.org/10.1016/S0168-8510(02)00115-X)
- Siegel, A. F. (2012). Histograms: Looking at the Distribution of Data. In *Practical business statistics* (6th ed., pp. 35–64). Academic Press. <https://doi.org/10.1016/B978-0-12-385208-3.00003-1>
- Smith, G. (2015). Probability. In *Essential Statistics, Regression, and Econometrics* (pp. 99–137). Academic Press. <https://doi.org/10.1016/B978-0-12-803459-0.00004-2>
- Srivastava, S., Kumar, P., Chauhan, S., & Banerjee, A. (2021). Household expenditure for immunization among children in India: A two-part model approach. *BMC Health Services Research*, 21(1), 1001. <https://doi.org/10.1186/s12913-021-07011-0>
- Statistics Netherlands. (2025a). *Spending on health care up by 8.1 percent in 2024*. <https://www.cbs.nl/en-gb/news/2025/22/spending-on-health-care-up-by-8-1-percent-in-2024>
- Statistics Netherlands. (2025b). *Spending on health and social care up by 8.9 percent in 2024*. <https://www.cbs.nl/en-gb/news/2025/41/spending-on-health-and-social-care-up-by-8-9-percent-in-2024>
- Tong, H. (2025). Least squares regression under weak moment conditions. *Journal of Computational and Applied Mathematics*, 458, 116336. <https://doi.org/10.1016/j.cam.2024.116336>
- Van de Ven, W., Van Kleef, R., & Van Vliet, R. (2013). Risicoselectie bij overstap zorgverzekeraar. *Economisch-Statistische Berichten*. https://repub.eur.nl/pub/50261/metis_193578.pdf

- Van Kleef, R. C., Van Vliet, R. C., & Van de Ven, W. P. (2013). 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., Bakx, P., Eijkenaar, F., Schut, F. T., Van de Ven, W. P. M. M., & Van Vliet, R. C. J. A. (2016). *Hoe problematisch is een onder- of overcompensatie? Position paper: bijdrage aan de fundamentele discussie over de risicoverevening*. <https://www.eur.nl/en/eshpm/media/26045workingpaperproblematischoverondercompensatie30>
- Van Kleef, R. C., McGuire, T. G., Van Vliet, R. C. J. A., & Van de Ven, W. P. M. M. (2017). Improving risk equalization with constrained regression. *The European Journal of Health Economics* 18, 1137–1156. <https://doi.org/10.1007/s10198-016-0859-1>
- Van Kleef, R. C., Schut, F. T., & Van De Ven, W. P. M. M. (2018). Premium Regulation, Risk Equalization, Risk Sharing, and Subsidies: Effects on Affordability and Efficiency. In *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets* (pp. 21–54). Elsevier. <https://doi.org/10.1016/B978-0-12-811325-7.00002-6>
- Van Kleef, R., & Van Vliet, R. (2025). *Eindrapport evaluatie risicoverevening 2024-2025*. <https://www.eur.nl/en/eshpm/media/2025-03-eindrapport-evaluatie-risicoverevening-2024-2025>
- Van Kleef, R. C., van Vliet, R., & Vanhommerig, J. (2026). *WPR-analyse risicoverevening 2026: verevenende werking en power*. Erasmus University Rotterdam. <https://www.eur.nl/eshpm/media/2026-05-eindrapport-wpr-analyse-rv26-v30apr2026-1>
- Vargas, V., & Wasem, J. (2006). Risk Adjustment and Primary Health Care in Chile. *Croatian Medical Journal*, 47(3), 459–468. [Risk adjustment and primary health care in Chile - PubMed](https://pubmed.ncbi.nlm.nih.gov/16111111/)
- Wammes, J. J. G., Tanke, M., Jonkers, W., Westert, G. P., Van Der Wees, P., & Jeurissen, P. P. (2017). Characteristics and healthcare utilisation patterns of high-cost beneficiaries in the Netherlands: A cross-sectional claims database study. *BMJ Open*, 7(11), e017775. <https://doi.org/10.1136/bmjopen-2017-017775>
- Wammes, J. J. G., Van Der Wees, P. J., Tanke, M. A. C., Westert, G. P., & Jeurissen, P. P. T. (2018). Systematic review of high-cost patients' characteristics and healthcare utilisation. *BMJ Open*, 8(9), e023113. <https://doi.org/10.1136/bmjopen-2018-023113>
- Werkgroep Ontwikkeling Risicoverevening. (n.d.). *Onderzoeksprogramma Risicoverevening 2025-2026 (WOR 1261)*. <https://www.rijksoverheid.nl/documenten/2025/09/22/wor-1261-onderzoeksprogramma-risicoverevening-2025-2026>
- Zorginstituut Nederland. (2019). *Het Zorgverzekeringsfonds (Zvf)*. <https://www.zorginstituutnederland.nl/financiering/fondsbeheer-zvf-en-flz/zorgverzekeringsfonds>
- Zorgverzekeringswet. (2026). <https://wetten.overheid.nl/BWBR0018450/2026-01-01>

7. Appendices

Appendix 1: Risk adjusters in the 2026 risk equalisation model

Risk adjusters that are used in the 2026 risk equalisation model for mental healthcare

Risk adjuster (abbreviation)	Risk classes	Description
Age/sex	35	Risk classes based on an interaction between age and sex
Pharmacy cost groups, mental disorder (<i>PCG</i>)	10	Morbidity indicator derived from the use of specific pharmaceuticals in the prior year
Diagnostic cost groups, mental disorder (<i>DCG</i>)	17	Morbidity indicator derived from specific hospital treatments in the prior year
Source of income, interacted with age (<i>SOI</i>)	43	Socioeconomic indicator based on the source of income (e.g., employed, self-employed, social security...)
GGZ-Region	10	Clusters of ZIP codes based on regional characteristics
Social Economic Status, interacted with age (<i>SES</i>)	10	Socioeconomic indicator based on the mean per-person household income
Persons per address, interacted with age (<i>PPA</i>)	18	Morbidity indicator based on whether the individual lives in a care facility and whether the person lives alone
Multiple year high costs (<i>MHC</i>)	8	Morbidity indicator based on whether the individual has had high healthcare costs (top X%) in at least one of the three previous years
Total number of risk classes	151	

Source: Ministerie van Volksgezondheid (2025b)

Risk adjusters that are used in the 2026 risk equalisation model for the mandatory deductible

Risk adjuster (abbreviation)	Risk classes	Description
Age/sex	30	Risk classes based on an interaction between age and sex
Source of income, interacted with age (<i>Source</i>)	29	Socioeconomic indicator based on the source of income (e.g., employed, self-employed, social security...)
Region	10	Clusters of ZIP codes based on regional characteristics
Multiple year high costs (<i>MHC</i>)	2	Morbidity indicator based on whether the individual has had high healthcare costs (top X%) in at least one of the three previous years
Total number of risk classes	71	

Source: Ministerie van Volksgezondheid (2025b)

Appendix 2: Standard evaluation subgroups

Subgroup		Description
Presence of at least one chronic condition		Enrollees who have a chronic condition which falls into the list of chronic conditions for evaluation ¹
Enrollees who have had pharmacy costs for multiple years		Enrollees who have had pharmacy costs above a certain cost threshold for multiple years.
Enrollees who have chosen the maximum voluntary deductible		Enrollees who have chosen an insurance contract which includes a voluntary maximum deductible
Standard agreement	Standard agreement with restrictive conditions	Enrollees who have chosen an insurance contract with restrictive conditions
	10% highest premiums	Enrollees who have an insurance contract which falls under the 10% highest premiums
	10% lowest premiums	Enrollees who have an insurance contract which falls under the 10% lowest premiums

Source: Ministerie van Volksgezondheid, Welzijn en Sport (2024)

¹The list of chronic conditions for evaluation is set out in Attachment 3 of the WOR-assessment framework.

Appendix 3: Overview of all additional variables in the dataset

Additional variable (abbreviation)	Description
Costs somatic care	The somatic healthcare costs of an individual in 2023.
Predicted costs somatic care	The predicted somatic healthcare costs of an individual for 2023, according to the 2026 risk equalisation model.
Insurance weight	The fraction of the year during which an individual was insured in 2023. This variable is used as a weight in each analysis .
15% Lowest costs	Highlights enrollees who, in 2020, belonged to the 15% of people with the lowest somatic healthcare costs.
15% Highest costs	Highlights enrollees who, in 2020, belonged to the 15% of people with the highest somatic healthcare costs.
Chronic illness	Highlights enrollees who, in 2022, had a chronic illness which is taken up in the chronic-illness-list by the Dutch Healthcare Institute.
Multiple Year Pharma Costs (MPC)	Risk adjuster which highlights enrollees who have had pharmacy costs for multiple years. This adjuster is used for constrained regression in the 2026 risk equalisation model.
Pharmacy Cost Group flag	Highlights enrollees with the presence of any PCG risk adjuster.
Diagnostic Cost Group flag	Highlights enrollees with the presence of any DCG risk adjuster.

Appendix 4: Stepwise regression; risk classes

Y1-regression	Y2-regression	P-regression
<i>Source of Income (SOI)</i>	<i>Pharmacy Cost Group (PCG)</i>	<i>Pharmacy Cost Group (PCG)</i>
A8: Disability benefit recipients: 18–34	Fkg1: Pharmacy Cost Group 1	Fkg1: Pharmacy Cost Group 1
A37: Reference group: 0–17	Fkg3: Pharmacy Cost Group 3	Fkg2: Pharmacy Cost Group 2
A40: Reference group: 45–54	Fkg4: Pharmacy Cost Group 4	Fkg3: Pharmacy Cost Group 3
<i>Socioeconomic Status (SES)</i>	Fkg5: Pharmacy Cost Group 5	Fkg4: Pharmacy Cost Group 4
S3: Very low SES: 70 years and older	Fkg7: Pharmacy Cost Group 7	Fkg5: Pharmacy Cost Group 5
S9: Middle SES: 70 years and older	Fkg8: Pharmacy Cost Group 8	Fkg7: Pharmacy Cost Group 7
S12: High SES: 70 years and older	Fkg9: Pharmacy Cost Group 9	Fkg10: Pharmacy Cost Group 10
<i>Region</i>	Fkg10: Pharmacy Cost Group 10	Fkg11: Pharmacy Cost Group 11
R2: Region cluster 2	Fkg11: Pharmacy Cost Group 11	Fkg13: Pharmacy Cost Group 13
R3: Region cluster 3	Fkg15: Pharmacy Cost Group 15	Fkg15: Pharmacy Cost Group 15
<i>Persons per Address (PPA)</i>	Fkg16: Pharmacy Cost Group 16	Fkg16: Pharmacy Cost Group 16
P10: WLZ institution without treatment or extramural WLZ, incoming age 18–69	Fkg17: Pharmacy Cost Group 17	Fkg17: Pharmacy Cost Group 17
<i>Age/Sex</i>	Fkg22: Pharmacy Cost Group 22	Fkg20: Pharmacy Cost Group 20
G15: Men, 60–64 years	Fkg26: Pharmacy Cost Group 26	Fkg22: Pharmacy Cost Group 22
<i>Historical Costs for Care Aids (HCA)</i>	Fkg27: Pharmacy Cost Group 27	Fkg26: Pharmacy Cost Group 26
Hk0: Cost group 0	Fkg28: Pharmacy Cost Group 28	Fkg27: Pharmacy Cost Group 27
<i>Multiple Year High Costs (MHC)</i>	Fkg31: Pharmacy Cost Group 31	Fkg28: Pharmacy Cost Group 28
M0: High-cost class 0	Fkg33: Pharmacy Cost Group 33	Fkg29: Pharmacy Cost Group 29
<i>Pharmacy Cost Group (PCG)</i>	Fkg35: Pharmacy Cost Group 35	Fkg33: Pharmacy Cost Group 33
Fkg0: No pharmacy cost group	Fkg36: Pharmacy Cost Group 36	Fkg37: Pharmacy Cost Group 37
	Fkg37: Pharmacy Cost Group 37	Fkg40: Pharmacy Cost Group 40
	Fkg41: Pharmacy Cost Group 41	Fkg41: Pharmacy Cost Group 41
	Fkg42: Pharmacy Cost Group 42	Fkg42: Pharmacy Cost Group 42
	Fkg43: Pharmacy Cost Group 43	Fkg43: Pharmacy Cost Group 43
	Fkg46: Pharmacy Cost Group 46	Fkg46: Pharmacy Cost Group 46
	<i>Age/Sex</i>	<i>Socioeconomic Status (SES)</i>
	G7: Men, 18–24 years	S4: Low SES, 0–17
	G8: Men, 25–29 years	S5: Low SES, 18–69
	G9: Men, 30–34 years	S7: Middle SES, 0–17
	G10: Men, 35–39 years	S8: Middle SES, 18–69
	G12: Men, 45–49 years	<i>Source of Income (SOI)</i>
	G13: Men, 50–54 years	A1: Reference group: 0–17
	G18: Men, 75–79 years	A5: Reference group: 55–64
	G19: Men, 80–84 years	A6: Reference group: 65–69
	G20: Men, 85–89 years	A11: Disability benefit recipients: 55–64
	G28: Women, 18–24 years	A12: Disability benefit recipients: 65–69
	G29: Women, 25–29 years	A17: Social assistance recipients: 55–64
	G30: Women, 30–34 years	A18: Social assistance recipients: 65–69
	G31: Women, 35–39 years	A19: Student 0–17
	G32: Women, 40–44 years	A20: Student 18–34
	G33: Women, 45–49 years	A25: Self-employed: 0–17
	G34: Women, 50–54 years	A26: Self-employed: 18–34
	G35: Women, 55–59 years	A27: Self-employed: 35–44
	G36: Women, 60–64 years	A28: Self-employed: 45–54
	G37: Women, 65–69 years	A29: Self-employed: 55–64
	G38: Women, 70–74 years	A30: Self-employed 65–69
	G39: Women, 75–79 years	A31: Highly educated: 0–17
	G40: Women, 80–84 years	A32: Highly educated: 18–34
	G41: Women, 85–89 years	A33: Highly educated: 35–44
	G42: Women, 90+ years	<i>Indicator Pregnancy (IP)</i>
	<i>Source of Income (SOI)</i>	B1: class 1
	A1: Reference group: 0–17	B2: class 2
	A4: Reference group: 45–54	<i>Age/Sex</i>
	A5: Reference group: 55–64	G3: Men, 1–4 years
	A8: Disability benefit recipients: 18–34	G12: Men, 45–49 years
	A9: Disability benefit recipients: 35–44	G13: Men, 50–54 years
	A10: Disability benefit recipients: 45–54	G20: Men, 85–89 years
	A11: Disability benefit recipients: 55–64	G21: Men, 90+ years
	A12: Disability benefit recipients: 65–69	G24: Women, 1–4 years
	A13: Social assistance recipients: 0–17	G35: Women, 55–59 years
	A15: Social assistance recipients: 35–44	G36: Women, 60–64 years
	A16: Social assistance recipients: 45–54	G37: Women, 65–69 years
	A17: Social assistance recipients: 55–64	G38: Women, 70–74 years
	A25: Self-employed: 0–17	G39: Women, 75–79 years
	A28: Self-employed: 45–54	G40: Women, 80–84 years
	A29: Self-employed: 55–64	<i>Region</i>
	A30: Self-employed: 65–69	R1: Region cluster 1
	A31: Highly educated: 0–17	R3: Region cluster 3

A40: Reference group: 45–54 A41: Reference group: 55–64 A42: Reference group: 65–69	R4: Region cluster 4 R5: Region cluster 5 R6: Region cluster 6 R7: Region cluster 7 R8: Region cluster 8 R9: Region cluster 9
<i>Region</i>	
R3: Region cluster 3 R4: Region cluster 4 R5: Region cluster 5 R6: Region cluster 6 R7: Region cluster 7 R8: Region cluster 8 R9: Region cluster 9 R10: Region cluster 10	<i>Diagnostic Cost Group (DCG)</i> Dkg1: Diagnostic Cost Group 1 Dkg2: Diagnostic Cost Group 2 Dkg18: Diagnostic Cost Group 18
<i>Indicator Pregnancy (IP)</i>	<i>Persons per Address (PPA)</i> P11: WLZ institution without treatment or extramural WLZ, incoming age 70–79 P3: WLZ institution with treatment, existing age 80+
B0: Class 0 B1: Class 1 B3: Class 3	<i>Historical Costs for Care Aids (HCA)</i> H0: No historical care aid costs
<i>Socioeconomic Status (SES)</i>	<i>Missing Costs History (MCH)</i> Kh4: Missing historical cost information, class 4
S2: Very low SES, 18–69 S5: Low SES, 18–69 S8: middle SES, 18–69	<i>Physical Therapy Diagnosis Group (PTD)</i> F3: Physical therapy diagnosis group 3
<i>Missing Costs History (MCH)</i>	
Kh5: Class 5 Kh1: Class 1 Kh4: Class 4	
<i>Persons per Address (PPA)</i>	
P5: WLZ institution without treatment or extramural WLZ, existing 70–79 P10: WLZ institution without treatment or extramural WLZ, incoming age 18–69	
<i>Historical Costs for Care Aids (HCA)</i>	
Hk4: Cost group 4	
<i>Multiple Year High Costs (MHC)</i>	
M1: High-cost class 1	
<i>Diagnostic Cost Group (DCG)</i>	
Dkg18: Diagnostic Cost Group 18	
<i>Multiple Year Costs Home Care (MCHC)</i>	
V8: Multiple-year home care cost group 8	

Based on the cost for the year 2023, and 17,736,376 observations. Individuals are weighted according to the proportion of the year for which they were insured. Additional information on the weighting factor is provided in Paragraph 3.2.