Combining the Dutch somatic and mental health care risk equalisation model into one model

Maartje van Tilburg – 726272
Erasmus University
Supervisors: R. Douven & J. Smeets
Rotterdam

18 June 2025

Word count: 9,672 words

Abstract

Risk equalisation plays a crucial role in ensuring fair competition among insurers and equitable access to health care in the Dutch system of regulated competition in the health insurance market, especially given open enrolment and community-rated premiums. Currently, two separate risk equalisation models are used – one for somatic care and one for mental health care. These models differ in predictive accuracy and require parallel development, increasing administrative complexity. Although somatic and mental health conditions are often interrelated, limited research has investigated whether combining these models could improve predictive performance. This study evaluates whether a combined model improves predictive accuracy by comparing it to the existing separate models using ordinary least squares (OLS) regressions. Data from the Central Bureau of Statistics were used, incorporating health care expenses from 2022 (t-3). Predictive accuracy was assessed using multiple indicators: R^2, Cummings Prediction Measure (CPM), Mean Absolute Percentage Error (MAPE), and subgroup undercompensation. Results indicate that the combined model performs similarly to the separate models on overall metrics, with a slight advantage in favour of the combined model. At the subgroup level, improvements are more pronounced, particularly for individuals with high mental health care expenses or chronic psychiatric illnesses, where the combined model demonstrates significantly higher accuracy. These findings suggest that a unified risk equalisation model could improve predictive performance, which would also be beneficial for the administrative burden. Further research is needed to examine how the combined model can be applied in practice.

Acknowledgment

I would like to take this opportunity to express my gratitude to those who supported me during the process of writing this thesis. First and foremost, I would like to thank my supervisors R. Douven and J. Smeets, for their guidance and insightful feedback throughout the research process. I am also grateful to the Ministry of Health, Welfare and Sport for offering me this internship and for the support and advice provided by many kind and helpful colleagues. Furthermore, I want to thank the Central Bureau of Statistics for granting me access to the data used in this study. Throughout the process, ChatGPT functioned as a digital brainstorming partner, and it assisted with language and grammar improvements.

I am grateful to the Netherlands Institute for Health Services Research (Nivel) for access to morbidity information registered by general practitioners. This study has been approved according to the governance code of Nivel Primary Care Database, under number NZR-00322.052. The use of electronic health records for research purposes is allowed under certain conditions. When these conditions are fulfilled, neither obtaining informed consent from patients nor approval by a medical ethics committee is obligatory for this type of observational studies containing no directly identifiable data (art. 24 GDPR Implementation Act jo art. 9.2 sub j GDPR).

Table of contents

Introduction

In the Netherlands, private health insurance for curative care is mandatory, with a community-rated premium for basic coverage and open enrolment under the Health Insurance Act (van Kleef et al., 2018). Community-rated premiums protect high-risk individuals from unaffordable high premiums but also incentivise health insurers to engage in risk-selection, defined as "actions by consumers and insurers to exploit unpriced risk heterogeneity" (van Kleef et al., 2024, p. 177). By attracting low-risk enrolees and deterring off high-risk enrolees, insurers seek to maximise profits and minimise losses. Risk selection has potential negative effects on solidarity between low-risk and high-risk individuals, and it could reduce efficiency and quality of care (van Kleef et al., 2013a).

To mitigate potential risk-selection, the Dutch government implemented an ex-ante risk equalisation model, which compensates insurers based on enrolees' risk profiles (van Kleef et al., 2018). A well-functioning risk equalisation system is therefore essential to reducing incentives for risk selection. The Netherlands uses two separate risk equalisation models: one for somatic care and another for mental health care. While both models share common risk-adjustors – such as sex, age, and source of income (AVI) (Ministerie van VWS, 2019) – they also include domain-specific risk-adjustors. Having two models results in a difference in predictive accuracy and two models that need maintenance. This raises the question of why expenditures for somatic and mental health care are estimated using two models instead of a single, integrated model – particularly given that a combined approach could improve predictive accuracy and reduce maintenance efforts.

The division of the risk equalisation models dates back to 2008, when mental health care was incorporated into the Health Insurance Act (van Kleef et al, 2018; Vektis, n.d.). At the time, mental health care had distinct risk-sharing mechanisms absent from somatic care, requiring separate cost predictions. Although these risk-sharing mechanisms no longer exist, the separation between somatic and mental health care persists. One possible reason for maintaining two separate models is the difference in distribution of health care costs: mental health care costs tend to be more right skewed than somatic care costs, making them more difficult to predict accurately. Another reason could be that the models have diverged so much over time that merging them now would pose significant challenges (van Kleef et al., 2018).

Developing a combined model is challenging because the existing models differ in risk-adjustors and predictive performance (van Kleef et al., 2018). For example, the somatic care model includes risk-adjustors such as Physiotherapy Diagnostic Groups (FDG) and Historic Somatic Morbidity (HSM), which are not present in the mental health care model. Predictive performance also varies: the somatic care model outperforms the mental health care model (Cattel et al., 2024b). However, when it comes to minimising risk selection, what matters most is the accuracy of predicting total health care expenses, not the separate components. Health insurers need adequate compensation for their enrolees' overall risk profile, regardless of whether this is achieved through a combined model or two separate models. Therefore, reducing incentives for risk selection can be accomplished either by maintaining two separate models or, as this study explores, by developing a single model that estimates the sum of somatic and mental health care expenses.

The goal of this research is to assess whether one risk equalisation model, that estimates the sum of somatic and mental health care expenditures, is an improvement over two separate risk equalisation models for somatic and mental health care. Using the following research question: *Will integrating the somatic and mental health care risk equalisation models into a single model result in more accurate predictions of total expected health care costs compared to keeping two separate models?*

Research on improving risk equalisation by combining somatic and mental health care models is limited. Van Asselt et al. (2015) investigated a combined risk equalisation model integrating the somatic care model with the short-term mental health care model, finding that the combined approach performed slightly better. However, since that time, the short-term and long-term mental health care models have been merged into a single model, and both the somatic and mental health care models have developed (Cattel et al., 2017). As a result, the impact of combining the updated somatic and mental health care models is uncertain. This uncertainty, alongside the potential for improved performance, underscores the importance of further research in this area.

This paper is structured as follows. Chapter 2 discusses the relevant theoretical background, including risk selection, the risk equalisation model, the possible outcomes when combining the two models, and relevant outcome measures. Chapter 3 includes the methodology and description of how analyses were conducted is presented. The results of the analyses are presented in Chapter 4, followed by the final discussion and conclusion in Chapter 5.

Theoretical framework

Risk selection

In the Netherlands, the health insurance market for curative care is characterised by regulated competition. This model encourages competition among both health care providers and health insurers (van Kleef et al., 2013a). Providers compete on quality and price to attract contracts from health insurers, while insurers compete with one another based on premiums and the quality of care they purchase from care providers (Enthoven & van de Ven, 2007). However, market competition can also lead to unfavourable outcomes. To mitigate these, the Dutch government enforces regulations aimed at maintaining a balance between solidarity and incentives for efficiency and quality improvement (van Kleef et al., 2013a). Solidarity is achieved through cross-subsidisation between low-risk individuals, such as the young and healthy, and high-risk individuals, such as the elderly or those with chronic illnesses. This requires effective risk pooling across a heterogeneous population. To ensure solidarity, the government has implemented several regulatory measures: mandatory health insurance for all residents, a standardised basic health care package that all insurers must offer, open enrolment policies for the basic health care package, and community-rated premiums, which prohibit insurers from varying premiums based on individual risk characteristics (van Kleef et al., 2013a).

While these regulations limit insurers' ability to engage in risk selection they also increase the incentive to do so, as risk selection becomes one of the few remaining strategies to achieve financial equivalence (van Kleef et al., 2013a). Risk selection can take several forms, including product differentiation, selective advertising and marketing, and one-stop shopping, where supplementary health insurance – exempt from open enrolment regulations – is used strategically to attract low-risk individuals (van Kleef et al., 2013a).

Risk selection is considered undesirable because it has three major consequences (van Kleef et al., 2013a). First, it undermines the principle of solidarity, as low-risk and high-risk individuals end up in different health plans, leading to premium differentiation based on health status rather than efficiency. Second, it reduces efficiency, as insurers may focus on identifying and attracting low-risk individuals instead of improving the overall efficiency of care delivery. Third, it diminishes the quality of care, as insurers may avoid investing in improvements that primarily benefit high-risk individuals, thereby discouraging their enrolment. To mitigate the incentives for risk selection, which are intensified by the very regulations designed to promote solidarity, the Dutch government has implemented a risk equalisation model for expected expenses within the basic care package (van Kleef et al., 2013a).

The risk equalisation model

The risk equalisation model mitigates potential incentives for risk selection by health insurers, which arise from unpriced risk heterogeneity – the variance between an insurance contract's predicted expenses and the income it generates (van Kleef et al., 2024). To address this, the Dutch government uses the risk equalisation model to predict annual health care expenses for each individual, based on risk-adjustors, such as age & sex or socioeconomic status (van de Ven et al., 2022). Based on these predictions, the government provides prospective payments to health insurers to compensate for the expected health care costs of their enrolees. After these payments are made, insurers bear the remaining financial risk. By reducing unpriced risk heterogeneity, the risk equalisation payments limit insurers' incentives to engage in risk selection (van Kleef et al., 2024).

The Dutch risk equalisation system is divided into a somatic and mental health care model (van Kleef et al., 2018). The third risk equalisation model for out-of-pocket expenses falls beyond the scope of this paper. An important distinction between the risk equalisation models for somatic and mental health care lies in the different risk-adjustors used (Cattel et al., 2024a). These risk-adjustors are organised into various classes. The somatic care model includes 12 risk-adjustors with a total of 225 classes. In contrast, the mental health care model includes 8 risk-adjustors with 130 classes (Cattel et al., 2024a). The complete list of risk-adjustors per model can be found in Table 1, along with the corresponding definitions of the abbreviations.

Table 1: Risk-adjustors per model (number of classes) and corresponding definition

Somatic care model	**Mental health care model**	**Definition**
Age & sex *(42)*	Age & sex *(30)*	Age & sex
AVI *(36)*	AVI *(29)*	Source of income
SES *(12)*	SES *(8)*	Socioeconomic status
PPA *(19)*	PPA *(18)*	Persons per address
FKG *(49)*	FKG *(10, different classes)*	Pharmacy cost groups
DKG *(27)*	DKG *(17, different classes)*	Diagnostic cost groups
Region *(10)*	Region *(10, different classes)*	Region
MHK *(9)*	MHK *(8, different classes)*	Multiple year high costs
FDG *(5)*		Physiotherapy diagnostic groups
MVV *(10)*		Multiple year V&V (nursing & care) costs
HSM *(2)*		Historic somatic morbidity
IBZ *(4)*		Indication pregnancy/delivery

Four risk-adjustors in Table 1, namely age & sex, AVI, SES and PPA, are shared by both the somatic and mental health care models (Cattel et al., 2024a). In addition, both models include FKG, DKG, region, and MHK, although the specific classes within these risk-adjustors differ between the two models. The somatic care model also incorporates four additional risk-adjustors that are absent from the mental health care model: FDG, MVV, HSM, and IBZ. Furthermore, the somatic care model uses MFK (multiple-year pharmacy costs) as a constraint in the regression (Cattel et al., 2024a). However, this constraint is not included in this study, as research on the risk equalisation model by the Ministry of Health, Welfare and Sport (VWS) is conducted without constraints (Smeets, 2024) and MFK was not included in the dataset. Therefore, MFK will not be discussed further. A final point of divergence between both models is that the mental health care model applies only to individuals aged eighteen and older, since mental health care expenses for minors are covered under the Youth Care Act rather than the Health Insurance Act (Nederlands Jeugdinstituut, n.d.)

The predictive performance of the mental health care model is lower than that of the somatic care model (van Asselt et al., 2015). This could be because of the model's limited predictive capacity or because mental health care expenses tend to be more right skewed, making them inherently more difficult to predict (van Asselt et al., 2015). As a result, ex-post regulations have been introduced to protect health insurers from excessive financial risk. These are regulations made after actual expenses are known, in contrast to the ex-ante payments, which are prospective payments made before care is delivered. For the somatic care model 75% of health care expenses above €400,000 are reimbursed (PWC, 2023; Cattel et al., 2024b). In contrast, the mental health care model is subject to both a bandwidth correction beyond -10/+10 euros (Agema, 2024) and 90% reimbursement for the top 0.5% expenses (Cattel et al., 2024b). However, a key drawback of ex-post regulations is that they weaken cost-containment incentives for insurers (Douven, 2010; PWC, 2023). Therefore, improving the predictive accuracy of the risk equalisation model is a priority, with the goal of phasing out reliance on ex-post measures – without increasing incentives for risk selection.

Combining the somatic and mental health care model

Combining the two models, by estimating the sum of somatic and mental health care expenses with one model, could be beneficial for the predictive accuracy. However, outcomes could also be unfavourable. From a theoretical standpoint, integration may improve the predictive accuracy, because of the strong interconnection between somatic and mental health care (Zheng et al., 2024). For example, patients with a depression have a 60% larger risk of developing type 2 diabetes, and patients with diabetes are two to three times more likely to develop a depression (Zheng et al., 2024). Health care costs for these types of patients are likely to be better predictable by a combined model, as it includes all risk categories and captures cross-domain correlations and interaction effects that separate models may miss.

Another reason why a combination of the two models might improve predictive accuracy is related to the skewness of cost distributions (van Asselt et al., 2015). Mental health care costs tend to be very right skewed, while somatic care costs are less skewed. Combining these costs could potentially reduce overall skewness if the less skewed somatic costs moderate the mental health care expenses (van Asselt et al., 2015). However, depending on the proportions and correlations, combining the costs might also increase overall skewness, making prediction of the health care expenses more challenging.

Empirical results also do not provide a definitive answer to the question of whether combining the somatic and mental health care model leads to improved predictive performance. For example, Germany's risk equalisation model demonstrates that both types of health care expenses can be predicted within a unified framework (Bundesministerium für Gesundheit, n.d.). While a 2013(b) study by van Kleef et al. compared the German and Dutch risk equalisation model – reporting a higher R^2 for the Dutch somatic care model – this comparison should be interpreted with caution, as differences in data and dependent variable limit the validity of a direct comparison, as will be explained later this chapter. Even if the Dutch somatic care model outperformed the German model, this does not necessarily imply that combining cost types reduces performance, since it is only compared with the somatic care model and not with the mental health care model. Moreover, there are more differences between the Dutch and German model than only the combination of health care costs. Lastly, both countries have since then updated and refined their models (van de Ven et al., 2024), making it even more difficult to determine whether combining health care costs leads to improved predictive accuracy.

Still, research by van Asselt et al. (2015) supports the combination of the two models. In their study, the short-term mental health care model was integrated with the somatic care model, resulting in a slight improvement in predictive performance. However, van Asselt et al.'s research focused only on short-term mental health care, whereas in 2018, the short-term and long-term models were combined following the transfer of long-term care to the Health Insurance Act in 2015 (Cattel et al., 2017). Although the models have evolved since then and the findings are therefore not conclusive for the current context, this research provides a valuable indication of the potential improved performance of combining the present-day somatic and mental health care models.

Outcome measures

To assess whether the predictive performance increases when combining the somatic and mental health care model multiple outcome measures are considered to make a balanced decision. One measure used in this study is the R^2 from the regression analysis – a measure that "indicates the proportion of the total variance in individual expenditures that is explained by the linear influence of the set of risk adjusters, and ranges between zero and one" (van de Ven & van Kleef, 2024, p. 5). In other words, it indicates how much of the variance between the values is explained by the variables in the regression. To guide this assessment, the study considers the work of van de Ven and van Kleef (2024), who discuss the R^2 as an indicator of how well risk equalisation payments reduce predictable profits and losses for insurers. They caution that the R^2 is difficult to interpret, can lead to wrong and misleading conclusions, and is not useful for measuring selection incentives. However, they also highlight that the R^2 is valuable for comparing different risk equalisation formulas when using the same dependent variable with the same dataset (van de Ven & van Kleef, 2024). Since this study uses the dependent variable for testing different models, R^2 will serve as a useful measure for analysing the predictive performance of the difference risk equalisation formulas.

Besides the R^2, Cumming's Prediction Measure (CPM) will also be used. CPM is "the proportion of the mean absolute deviation from the mean expenses that is explained or predicted by the linear influence of the set of risk adjusters" (van de Ven & van Kleef, 2024, p. 12). This means that the

CPM, just as the R^2, can describe the variance explained by the model on a scale from zero to one (Cumming et al., 2002). The R^2, however, is very sensitive to large errors, because the square of each prediction error is taken, while CPM uses the absolute values of these errors, and is therefore less sensitive to these large outliers (Cumming et al., 2002).

Lastly, the Mean Absolute Percentage Error (MAPE) will be taken as an outcome measure. The MAPE is the average of the absolute percentage errors (Kim & Kim, 2016). It is calculated by dividing the residuals (i.e. actual expenses minus predicted expenses) by the actual expenses and multiplying by 100. MAPE is especially useful to gain insight in relative errors, since it provides an intuitive measure (de Myttenaere et al., 2016). The lower the MAPE the better, since a MAPE of 88% implies that on average the predictions are 88% off from the real values (Coralogix, 2023).

Other outcome measures are based on the criteria established by the Ministry of VWS (Cluster Risicoverevening, 2024). These measures are divided into individual-level and subgroup-level indicators. Individual-level indicators show how well the risk equalisation works per person, while the subgroup-level indicators assess the predictive accuracy within certain groups in the total population. The subgroup level is considered more significant, as health insurers primarily perform risk selection at this level. In this study both levels will be evaluated. The R^2, CPM and MAPE will be used to evaluate the individual level. For the subgroup level the under- and overcompensation for different subgroups (e.g. yes/no mental health care costs) will be assessed.

Research methods

Data collection

For this study, data from the Central Bureau of Statistics (CBS) were used. Access to the data for scientific research purposes was requested and granted. The CBS provides data on all health care expenses and risk-adjustors per individual in the Netherlands for a specific year. A sample of 1.6 million individuals, for whom data from Nivel Primary Care Database (Nivel-PCD) were available, was used for analysis. Nivel-PCD contains data from general practitioners regarding health care usage and diagnoses (Vanhommerig et al., 2025; Nivel, 2022; Overbeek, n.d.). As a result, additional information on both somatic and mental health conditions is available for all individuals, making it possible to use this information for accurate subgroup analyses. All analyses were conducted using Stata 16.0, and Excel 2016 was used to visualise the graphs presented.

For this sample, data on somatic and mental health care expenses for the year 2022 (t-3) were available, along with dummy variables for the classes of the risk-adjustors in 2024 (t-1), corresponding with the required risk-adjustors of the 2025 risk equalisation model (Table 1). All individuals under the age of eighteen were excluded from the analysis, as their mental health care expenses were not included in the data due to these being reimbursed under the Youth Act rather than the Health Insurance Act. Classes relevant only to individuals under the age of eighteen, such as socioeconomic status for the ages zero to seventeen, were also excluded. Health care expenses were adjusted to reflect full-year enrolment. To account for this, the dataset included a weight factor that corrected for both the duration of enrolment and for how representative each individual was of the total population covered by the Health Insurance Act. Since individuals in the Nivel-PCD data were not fully representative of the total population, those who are underrepresented received a higher weight, while those who are overrepresented received a lower weight – it is the inverse of the probability of an observation being selected into the sample.

Since all classes of the risk-adjustors were dummy variables, a reference group was defined for the analysis. The reference group is defined as men between 18 and 24 years old, with a very low socio-economic status, who are employed, not living in a long term care facility or a one person household (residual group of persons per address), have no DKG, FKG or MHK for somatic or mental health illnesses, live in region 1 for the somatic care model and region 1 for the mental health care model[1], and have no FDG, MVV, HSM or IBZ.

Primary analysis: regression

In this study, two models are discussed and compared. The combined model estimated total health care costs using all available risk-adjustors, while the separate models consisted of distinct models for somatic and mental health care costs. To assess whether total health care expenses can be predicted

[1] Region is not a specific location but is formed by taking into account environmental factors (e.g. number of psychiatric facilities in the neighbourhood)

more accurately using a combined model than with separate models, both the combined and separate models were estimated using Ordinary Least Squares (OLS) regression. In practice, since 2024, the somatic care model has used a constrained regression approach instead of OLS (Cattel et al., 2024a). However, all preliminary research on the somatic risk equalisation model is still conducted using unconstrained OLS regression (Smeets, 2024). The mental health care model, in contrast, applies quadratic programming in practice to ensure that the normative cost assigned to each insured individual is greater than or equal to zero (Cattel et al., 2024a). Although these methods offer more precise control, OLS provides a sufficiently accurate and widely accepted baseline for comparing model performance in this study. While this simplification does not fully capture the complexity of the operational models, it enables a transparent and interpretable comparison, which is the aim of this study.

The following formulas were used to estimate the separate models:

$$y_{somatic} = \alpha X_{somatic} + \varepsilon_{somatic}$$

$$y_{mental\ health} = \beta X_{mental\ health} + \varepsilon_{mental\ health}$$

These models were estimated using weighted linear regression, where the health care expenses (y) were regressed on risk-adjustor classes in X with corresponding coefficients α and β, and residuals ε. Probability weights (pweight) were applied to ensure that individuals with partial-year enrolment and individuals that were over- or underrepresented were appropriately weighted in the estimation. Pweight is a function in Stata, used to account for unequal probabilities of selection in survey data. If an individual has a higher weight, this results in a higher contribution to the estimation of the regression coefficients, but the variables do not get multiplied by the weight (Linacre, n.d.; StataCorp, 2019). The variable $X_{somatic}$ included the risk-adjustor classes for the somatic care model, while $X_{mental\ health}$ included the risk-adjustor classes for the mental health model, as described in Table 1.

For the combined model, total health care expenses ($y_{total} = y_{somatic} + y_{mental\ health}$) were estimated using the following formula:

$$y_{total} = \gamma X_{total} + \varepsilon_{total}$$

In this formula, the total predicted health care costs ($\hat{y}_{total}$) were obtained by multiplying the estimated coefficient for each variable (γ) by the corresponding risk-adjustor class (X_{total}), resulting in the error term ε_{total}. The variable X_{total} consists of all risk-adjustor classes from both the somatic and mental health care models, with any duplicates included only once. As with the separate models, the regression was estimated using probability weights (pweight) to account for differences in enrolment duration and representation for the population enrolled under the Health Insurance Act.

Primary analysis: comparison

After estimating both the combined and separate models, the results were compared to evaluate predictive accuracy. The primary objective was to compare the predicted total health care costs from the combined model ($\hat{y}_{\text{total}}$) with the sum of the predictions from the separate models ($\hat{y}_{\text{somatic}} + \hat{y}_{\text{mental health}}$). The predictive performance of models was assessed using R^2, CPM and the MAPE. In addition, the individual R^2 values for the somatic and mental health care model were calculated, as well as the R^2 values obtained by using all risk-adjustors from the combined model to estimate somatic and mental health care costs separately.

Initially, only the classes of one risk-adjustor – age & sex – was included in the models, and the performance metrics were reported. Subsequently, additional risk-adjustors were added incrementally, and the analysis was repeated at each step. The risk-adjustors were added in the following order: age & sex, SES, AVI, PPA, DKG, FKG, MHK, region, FDG, MVV, IBZ, and HSM. The process began with the risk-adjustors common to both models, followed by those that differed in classes, and finally, those unique to the somatic care model.

The comparison between the R^2 values was done using the following formulas:

$$R^2_{total} = 1 - \frac{\sum_{i=1}^{n}(\varepsilon_{i,total})^2}{\sum_{i=1}^{n}(y_{i,total} - \bar{y}_{total})^2} \quad vs. \quad R^2_{apart} = 1 - \frac{\sum_{i=1}^{n}(\varepsilon_{i,somatic} + \varepsilon_{i,mental\ health})^2}{\sum_{i=1}^{n}(y_{i,somatic} + y_{i,mental\ health} - \bar{y}_{somatic} - \bar{y}_{mental\ health})^2}$$

Here, $\bar{y}_{\text{total}}$ denotes the mean of total health care expenses, while $\bar{y}_{\text{somatic}}$ and $\bar{y}_{\text{mental health}}$ represent the mean of somatic and mental health care costs, respectively. Pweights was incorporated in the analysis as described earlier. It is important to note that $\varepsilon_{\text{total}}$ does not necessarily equal $\varepsilon_{\text{somatic}} + \varepsilon_{\text{mental health}}$, due to the separate estimation processes and different coefficient structures. As a result, the R^2 value of the combined model can differ from that of the separate models, which is an indicator of predictive performance.

The comparison between the CPM and MAPE was conducted using the following two formulas:
For CPM:

$$CPM_{total} = 1 - \frac{\sum_{i=1}^{n}|\varepsilon_{i,total}|}{\sum_{i=1}^{n}|y_{i,total} - \bar{y}_{total}|} \text{ vs. } CPM_{apart} = 1 - \frac{\sum_{i=1}^{n}|\varepsilon_{i,somatic} + \varepsilon_{i,mental\ health}|}{\sum_{i=1}^{n}|y_{i,somatic} + y_{i,mental\ health} - \bar{y}_{somatic} - \bar{y}_{mental\ health}|}$$

For the MAPE:

$$MAPE_{total} = 100\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\varepsilon_{i,total}}{y_{i,total}}\right| \text{ vs. } MAPE_{apart} = 100\frac{1}{n}\sum_{i=1}^{n}\left|\frac{\varepsilon_{i,somatic} + \varepsilon_{i,mental\ health}}{y_{i,somatic} + y_{i,mental\ health}}\right|$$

From this point onward, all analyses included the full set of risk-adjustors. To assess whether the outcomes differed significantly, bootstrap analyses were performed. Bootstrapping involves repeatedly drawing random samples from the data with replacement, meaning some observations may be included more than once, while others may not be included at all (Stat20, n.d.). This process was repeated 500 times for all analyses, which is the recommended number for obtaining a confidence interval (Stat20, n.d.). In each iteration, both the R^2 for the combined model and that of the separate models were calculated using the same bootstrap sample. The difference between the two R^2 values was then computed by subtracting the R^2 of the separate models from that of the combined model (R^2_{total} - R^2_{apart}). After the bootstrap procedure, the mean difference, 95% confidence interval, standard error, and p-value were reported by Stata. This entire bootstrap analysis was conducted separately for CPM and the MAPE as well.

In this study, eight classes were omitted from the regression due to collinearity. These were primarily age-specific classes, such as those related to age and sex, AVI, SES, and PPA. Because the bootstrap analysis could not be performed with this many omitted variables, the excluded classes were progressively combined with adjacent or related age classes within the same risk-adjustor. This iterative process continued until only one class remained omitted. To maintain model interpretability and avoid excessive aggregation, a maximum of three classes were combined per grouping. This approach resulted in 35 original classes being consolidated into 15 new classes. The complete list of class-combinations is provided in Appendix 1. Only one class, FKG2 of the mental health care model (addiction excluding nicotine addiction), remained omitted only in the combined model. This class could not be meaningfully merged with any other class and was therefore not combined. To assess the impact of combining classes, the R^2, CPM and MAPE were also calculated for the regression using the combined classes. All analyses, except for the bootstraps, were conducted using the model without combined classes.

Primary analysis: subgroup analysis

For the subgroup analyses, the over- and undercompensation within several subgroups was assessed. The following subgroups were examined: the top 5% of individuals based on somatic care expenses, the top 5% of individuals based on mental health care expenses, the top 5% of individuals based on total expenses, individuals with chronic somatic illnesses (including all 106 chronic somatic illnesses reported by the Nivel-PCD data), individuals without chronic somatic illnesses, individuals with chronic psychiatric illnesses (including five chronic psychiatric illnesses reported by the Nivel-PCD data), and individuals without chronic psychiatric illnesses. Over- and undercompensation were assessed by calculating the weighted mean residuals for each subgroup. Individual residuals (i.e. the difference between actual and predicted expenses) were multiplied by the corresponding individual weights. The average of these weighted residuals was then divided by the mean weight of 11.85 to prevent inflation of the residuals due to the weighting. Paired t-tests were used to assess whether the predicted costs from the combined and separate models differed significantly per subgroup.

Additional analyses

In addition to the main analyses, additional analyses were conducted to provide further insight into the relationship between somatic and mental health care expenses. First, the relation between somatic and mental health care expenses was examined by calculating the correlation between the two. This correlation informs the potential value of combining them into a single model. In the case of a strong negative correlation, high somatic care costs tend to be associated with low mental health care costs, and vice versa. In this case, the combined model would need to capture two opposing trends, likely resulting in reduced predictive power. However, a negative correlation can also help reduce the right skewness of the total health care costs, as extreme values in one domain are offset by lower values in the other. Conversely, a strong positive correlation implies that both types of costs tend to increase or decrease together. In that case, combining the models might not add much predictive value, as the shared variance is already captured. Additionally, the overall skewness of total costs may increase, since high costs from both domains are compounded. If the correlation is close to zero, this suggests that somatic and mental health care expenses are largely independent. A combined model may then improve predictive performance by capturing additional variance associated with each domain, which is not explained by the other.

Second, to assess the effect of excluding individuals under the age of eighteen, the somatic care model was re-estimated, including all individuals and associated classes for this age group. The R^2 of this revised model was compared to the model used in this study and to that of the real model, providing insight into how representative this simplified research is for the real Dutch risk equalisation model.

Results

Data description

The final dataset included 1,614,109 individuals, of whom 1,184,748 were above the age of eighteen, and thus included. When applying weights, this corresponded to 17,310,265 individuals in the total population, with 14,044,432 individuals aged eighteen or older. To give a visual representation of the individuals included, regarding their sex and age, source of income and socioeconomic status, their frequencies are shown in Figure 1, 2 and 3. The numbers are weighted, since all analyses were conducted using the weight correction.

Figure 1
Number of weighted individuals per age and sex class

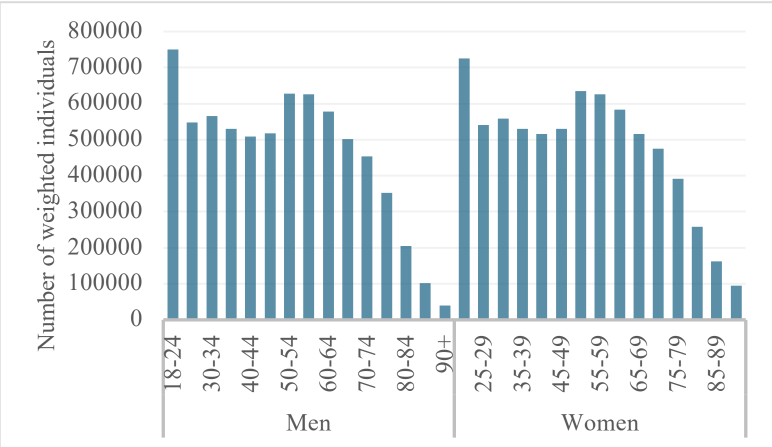


Figure 2
Number of weighted individuals per socioeconomic status class

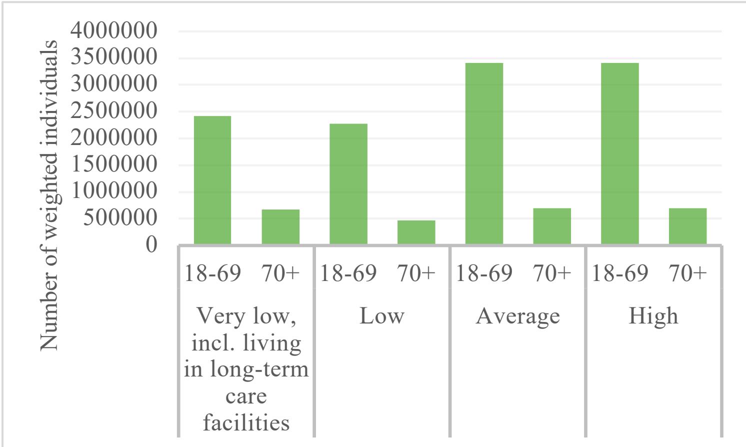


Figure 3
Number of weighted individuals per source of income class

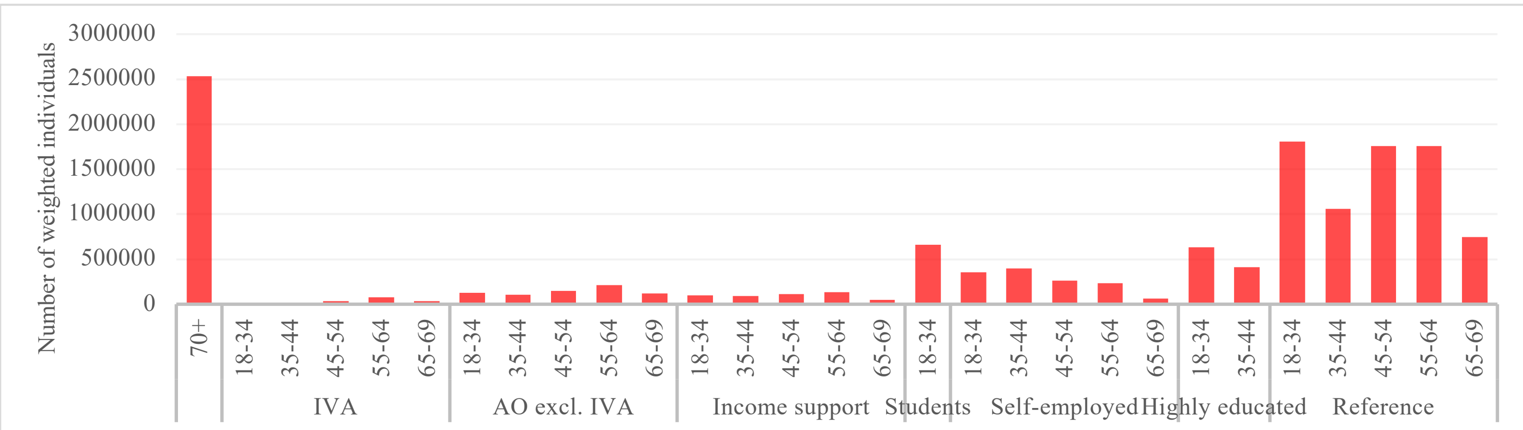


Figure 1 shows the number of weighted individuals on the y-axis and on the x-axis the different age groups for men and women. As can be seen, the distribution of men and women was nearly equal, with a slightly higher proportion of women included, mostly in the older age groups. Figure 2 shows the number of weighted individuals per class within the socioeconomic status risk-adjustor. Socioeconomic status was divided into four groups: very low, low, average and high. These groups were further divided by age, with categories for individuals below and above 70 years old. In the graph can be seen that there were fewer individuals with low socioeconomic status compared to individuals with very low, average and high socioeconomic statuses, which were almost equally distributed.

The graph in Figure 3 shows the number of weighted individuals per source of income risk-adjustor class. Source of income was divided in eight groups, which were further divided by age. The IVA group consists of individuals who were incapacitated for work with little to no prospect of recovery. The group AO excl. IVA included individuals who were also unable to work, but did not meet the criteria for IVA. The remaining categories comprised recipients of income support, students, self-employed individuals, highly educated individuals, individuals above 70 years old (retired) and the reference group, which included those employed by an organisation or company. This reference group was the largest group across all age categories, whereas the IVA group had a minority for all age groups.

The distribution of the health care expenses for both somatic and mental health care are shown in Figure 4 and 5, with the percentage of the total population on the y-axis and the amount of health care expenses in euros on the x-axis. Both somatic and mental health care costs were severely right skewed. For somatic care costs 74.0% of the individuals had costs between €0 and €2,000, and 2.7% had expenses above €20,000. For mental health care costs 93.7% of the individuals had costs between €0 and €100 and 4.7% of the individuals had expenses above €1,000. Next to the total distribution, also the distribution of health care expenses was shown for both the classes of age and sex and source of income. In Figure 6 the health care costs per age and sex class is shown. It shows that somatic care costs were increasing with age, while mental health care costs decreased. Somatic care costs for women were higher until the age of 55-59, afterwards costs for men were higher.

Figure 4
Distribution of somatic care costs

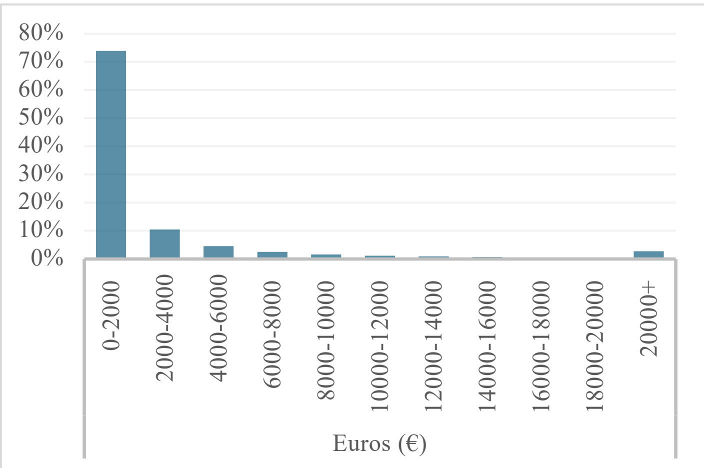


Figure 5
Distribution of mental health care costs

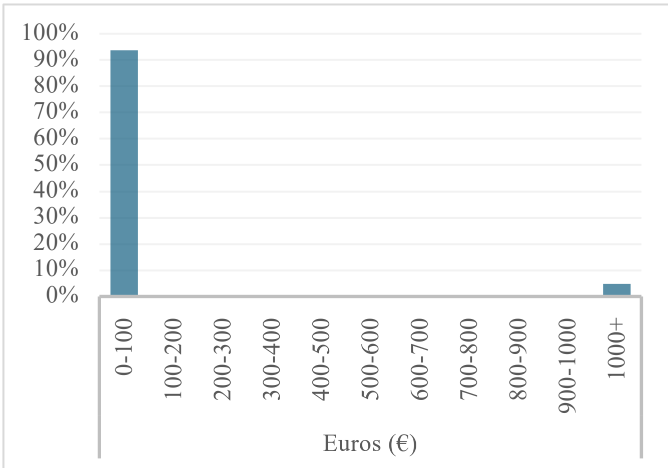


Figure 6
Mean individual somatic and mental health care costs per age and sex class

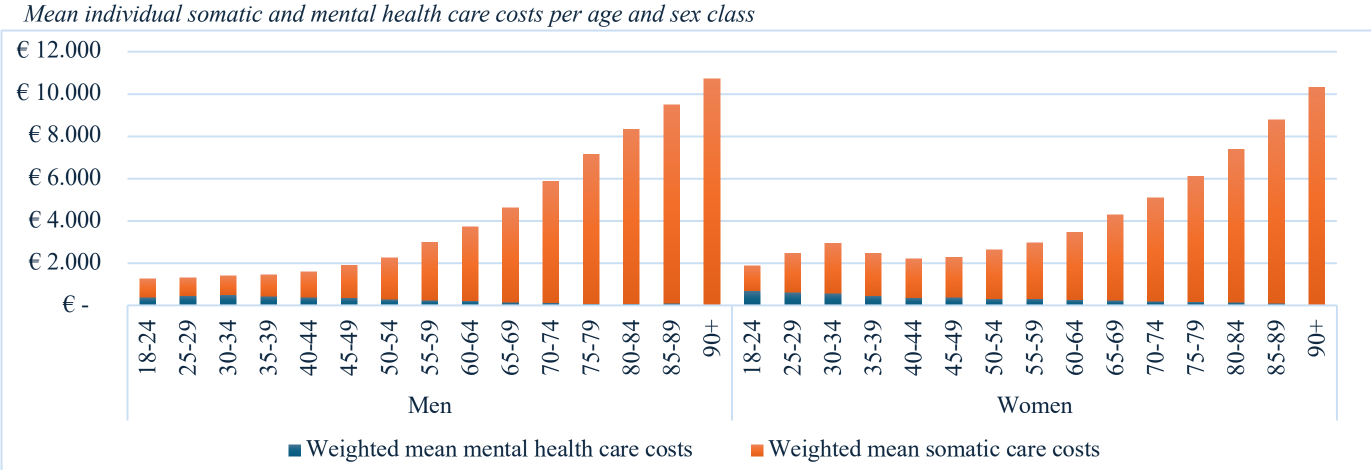


Figure 7
Mean individual somatic and mental health care costs per source of income class

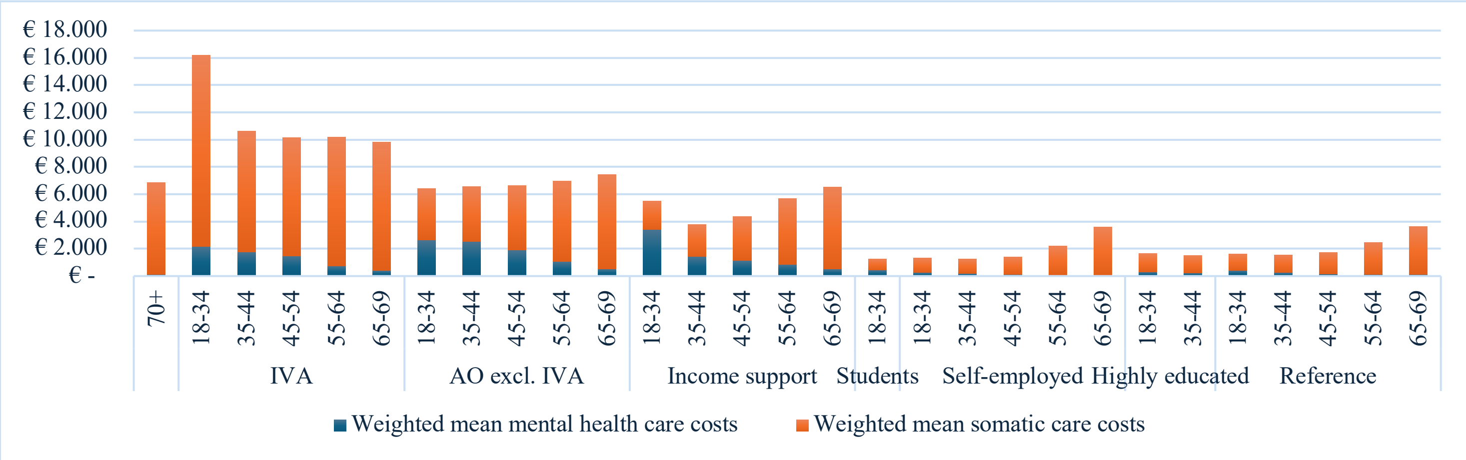


Figure 7 shows the health care expenses on the y-axis and the source of income classes on the x-axis. Illustrating that the health care expenses differed between these classes. Individuals that were in the IVA, AO excl. IVA or income support classes, had higher health care expenses compared to the other classes, for both somatic and mental health care expenses. Individuals in the 70+ class also had higher health care expenses, but these consisted of almost only somatic care costs.

Primary analysis: regression & comparison

The results of incrementally adding new risk-adjustors to both models are shown in Figures 8, 9, and 10. On the y-axis the values of the corresponding outcome are shown, while the x-axis shows which risk-adjustor is added onto the previous ones in the model. The R^2 is shown in Figure 8, where it increased with every risk-adjustor added for both the combined and separate models. At all points, the R^2 of the combined model was either equal or improved compared to the separate models. As was also the case for the CPM, which is shown in Figure 9. Initially, R2 was lower than CPM, but this relationship inversed after the addition of the DKG risk-adjustor. Only when all risk-adjustors were included, and solely in the combined model, did the CPM again exceed the R^2. The MAPE, shown in Figure 10, increased after adding the second and third risk-adjustor, but after that it decreased. A lower MAPE was better than a high MAPE. For all points the MAPE was either equal or lower for the combined model.

Figure 8
R^2 per added risk-adjustor

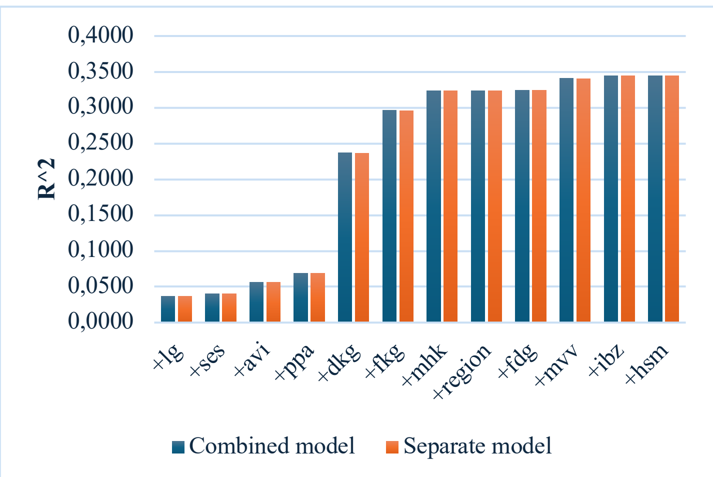


Figure 9
CPM per added risk-adjustor

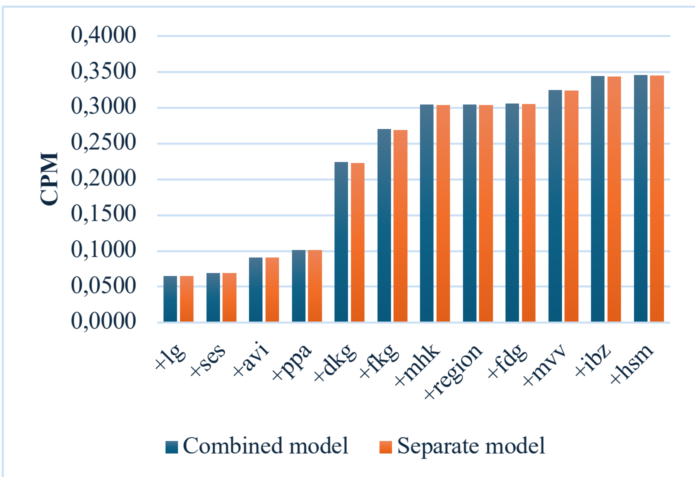
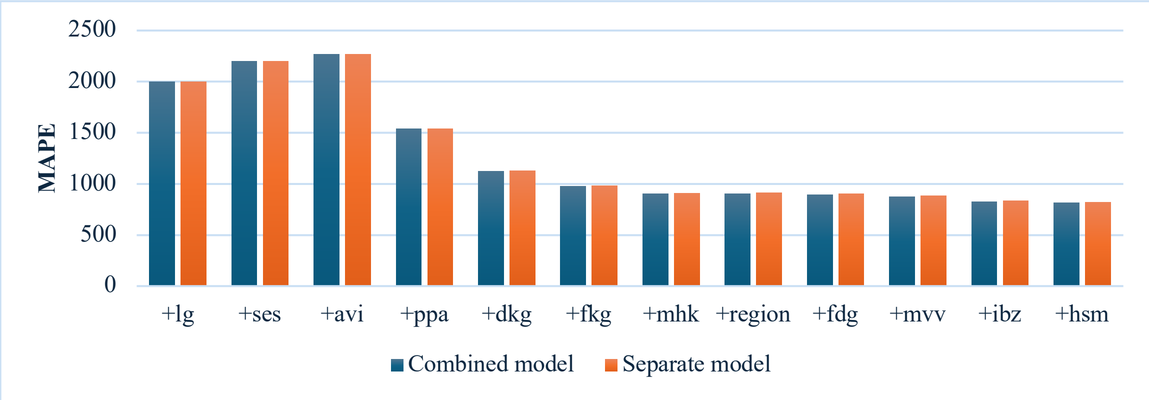


Figure 10
MAPE per added risk-adjustor



The difference between the combined model and the separate model for each added risk-adjustor is shown in Table 2. All values were in favour of the combined model. For the R^2 and CPM the values were positive, which means the combined model had a higher R^2 and CPM. For the MAPE the values were negative, which means that the MAPE was lower for the combined model compared to the separate model, and thus better. As can be seen, the differences between the two models only occurred after the DKG has been added, before then, the difference between all three outcome measures was zero.

Table 2
Difference in outcome measures between the combined and separate models per risk-adjustor added

	R^2	CPM	MAPE
+lg	0.0000	0.0000	0.0
+ses	0.0000	0.0000	0.0
+avi	0.0000	0.0000	0.0
+ppa	0.0000	0.0000	0.0
+dkg	0.0003	0.0016	-8.0
+fkg	0.0003	0.0010	-4.3
+mhk	0.0002	0.0009	-9.2
+region	0.0002	0.0009	-10.0
+fdg	0.0002	0.0009	-9.4
+mvv	0.0002	0.0010	-8.7
+ibz	0.0002	0.0009	-7.2
+hsm	0.0002	0.0008	-5.5

Both the somatic and mental health care model were included in the separate models. Including all risk-adjustors, the somatic care model achieved an R^2 of 0.3530, while the mental health care model had an R^2 of 0.2565. These results yielded an R^2 of 0.3447 for the separate models. In contrast, the combined model – after including all variables – achieved an R^2 of 0.3449. When estimating only somatic care costs, using all risk-adjustors of the combined model, the R^2 was 0.3531; for mental health care costs the R^2 was 0.2575. The correlation between the somatic and mental health care expenses was low, at 0.018 (Table 4). The CPM was 0.3452 for the combined model and 0.3444 for the separate models. Finally, the combined model had a MAPE of 816, while the separate models had a MAPE of 822.

To run the bootstrap, 35 classes were combined into 15 combined classes. After the classes were combined, the R^2 was 0.3446 and 0.3444 for the combined and separate models, respectively. This was a change of 0.0003 for both the combined and separate models compared to the models without combined classes. The CPM differed with 0.0001 from the original models after classes were combined, with a CPM of 0.3451 for the combined model and 0.3443 for the separate models. The MAPE was 806 and 812 for the combined and separate models, respectively, changing both with 10 compared to the original models. The difference between the combined and separate models, as calculated in the bootstrap, was 0.00019 for the R^2, in favour for the combined model (95% CI: 0.00013 – 0.00026). This difference was statistically significant with a p-value of 0.000. For the CPM the difference was 0.00078 (95% CI: 0.00057 – 0.0010), which was also in favour of the combined model and statistically significant (p-value = 0.000). Lastly, the MAPE differed 5.256 (95% CI: 0.622 – 9.889) between the combined and separate models, in favour of the combined model, and was also statistically significant with a p-value of 0.026.

Primary analysis: subgroup analysis

Per subgroup the mean undercompensation was calculated, to assess the predictive accuracy. For all subgroups, the combined model has less over- or undercompensation compared to the separate models. The mean undercompensation per subgroup is presented in Table 3. It shows the number of individuals and the mean weighted undercompensation per subgroup for both the combined and the separate models. The top 5% of individuals based on somatic care costs included 59,238 non-weighted individuals. In both the combined and the separate models, these individuals were severely undercompensated with an average of €15,811 and €15,819, respectively. The difference between the models was only €9 but it was statistically significant with a p-value of 0.000. At this subgroup level, there was a very small correlation between somatic and mental health care expenses, with a coefficient of 0.053. Table 4 shows all correlation coefficients per subgroup.

The top 5% of individuals based on mental health care costs included 59,237 non-weighted individuals, who were less undercompensated compared to those in the top 5% of individuals based on somatic care expenses. However, they still experienced a mean undercompensation of €3,600 with the combined model and €3,749 with the separate models. The €149 difference, in favour of the combined model, was also statistically significant (p-value = 0.000). At the sample level, this corresponds to a decrease of €8.8 million in undercompensation for this subgroup. The correlation between the somatic and mental health care expenses in this subgroup had a coefficient of -0.006.

For individuals in the top 5% of individuals based on total expenses, the combined model resulted in €48 less undercompensation compared to the separate models, which was statistically significant with a p-value of 0.000. Still, the mean undercompensation was €17,539 and €17,587 per individual for the combined and separate models, respectively. For this subgroup, the correlation coefficient was the largest, compared to the other subgroups, with a correlation coefficient of -0.131.

For the subgroups with and without chronic somatic or psychiatric illnesses, the results also favour the combined model. Individuals with a chronic psychiatric illness (n = 34,313) have on average €73 less undercompensation with the combined model compared to the separate models (€51 vs. €123, respectively). This difference was statistically significant (p-value = 0.000). Among the 1,150,435 individuals without a chronic psychiatric illness, the combined model resulted in €2 less overcompensation compared to the separate models, which was also statistically significant (p-value = 0.000). At the sample level, this corresponded to a reduction of slightly more than €2 million in overcompensation. The correlation coefficient for these subgroups was 0.027 for individuals with a chronic psychiatric illness and 0.014 for those without.

The last two subgroups, individuals with or without chronic somatic illness, had a correlation coefficient of 0.016 and 0.023, respectively. Individuals with a chronic somatic illness (n = 631,225) had a mean undercompensation of €56 with the combined model and €60 for the separate models. The difference of €4 was statistically significant (p-value = 0.000). The 553,532 individuals without chronic somatic illness were €5 less overcompensated using the combined model compared to the separate models (€64 vs. €69, respectively). This difference was statistically significant with a p-value of 0.000.

Table 3
Mean undercompensation per individual per subgroup

Subgroup	Individuals per subgroup	Mean undercompensation combined model	Mean undercompensation separate models	Difference in undercompensation between combined and separate model
Top 5% of individuals based on somatic care costs	59,238	€ 15,811	€ 15,819	-€ 9
Top 5% of individuals based on mental health care costs	59,237	€ 3,600	€ 3,749	-€ 149
Top 5% of individuals based on total health care costs	59,237	€ 17,539	€ 17,587	-€ 48
Chronic psychiatric illness	34,313	€ 51	€ 123	-€ 73
No chronic psychiatric illness	1,150,435	-€ 2	-€ 4	€ 2
Chronic somatic illness	631,225	€ 56	€ 60	-€ 4
No chronic somatic illness	553,532	-€ 64	-€ 69	€ 5

Additional analyses

The R^2, including also individuals under the age of eighteen, for the somatic risk equalisation model was 0.3198.

Table 4

Correlation coefficient per subgroup for somatic care costs v. mental health care costs

Group	***Correlation coefficient***
Total cohort	0,018
Top 5% of individuals based on mental health care costs	0,053
Top 5% of individuals based on somatic care costs	-0,006
Top 5% of individuals based on total health care costs	-0,131
Chronic psychiatric illness	0,027
No chronic psychiatric illness	0,014
Chronic somatic illness	0,016
No chronic somatic illness	0,023

Discussion and conclusion

Summary of main key findings

This study set out to examine the effects of combining the Dutch somatic and mental health care risk equalisation models compared to maintaining two separate models. The central research question was: *Will integrating the somatic and mental health care risk equalisation models into a single model result in more accurate predictions of total expected health care costs compared to keeping two separate models?* The results indicate that, for every outcome measure, the combined model improves predictive performance relative to the separate models. Statistically significant improvements were observed in the R2, CPM and MAPE metrics, with differences of 0.00019, 0,00078, and 5.256 respectively. At the subgroup level, the combined model showed more notable effects. In almost all subgroups, the mean undercompensation was lower using the combined model, indicating better alignment between predicted and actual costs. The only exception was among individuals without chronic somatic or psychiatric conditions, where overcompensation decreased.

Interpretation of findings

To assess the practical relevance of the findings, it is important to look beyond the statistical significance. Although the improvements in R^2, CPM and MAPE are statistically significant, the absolute differences are minimal. For instance, a change of 0.00019 in R^2 is unlikely to be meaningful on an individual level. The typical monotonicity property of R^2 – that it will always increase when additional variables are included (James et al., 2021) – does not apply here, since this analysis compares total costs in the combined model to the sum of predictions from the two separate models, rather than adding predictors to a single model. As such, these findings provide an important insight: predictive performance does not decline when switching to a combined model, which supports the feasibility of integrating the somatic and mental health care model.

The relevance of the differences in over- and undercompensation at the subgroup level must be interpreted considering subgroup sizes. A relatively large reduction in undercompensation for a small subgroup may have less overall impact than a modest reduction for a large subgroup. As shown in the results, the most substantial improvement is observed individuals with the top 5% of individuals based on mental health care costs. This group contains almost all individuals with mental health care expenses, since 94% of the individuals have no mental health care costs. For this group, the decrease in undercompensation corresponds to an average per-person improvement of €149, which translates into an estimated total effect of €8.8 million on the sample level (unweighted). If health insurers engage in risk selection, they are more likely to do so at the group level, as such groups are easier to identify. Therefore, the substantial decrease in undercompensation for individuals with high mental health care expenses represents a meaningful improvement. This reduction could significantly decrease insurers' incentives to avoid individuals with high mental health care costs. Taken together, these findings suggest that a combined risk equalisation model could be beneficial, particularly in improving health care expenses for high-cost mental health users, without introducing distortions in other subgroups.

The improvement in predictive accuracy observed with the combined model is more likely attributable to the reduction in skewness than to the mere addition of risk-adjustors. When comparing the R^2 of the somatic and mental health care models to those obtained by separately estimating somatic and mental health care costs using the combined model's risk-adjustors, the increase in R^2 is 0.0001 for somatic care costs and 0.001 for mental health care costs. This modest gain suggests that simply adding more risk-adjustors, particularly in the case of mental health care expenses, does not substantially improve predictive accuracy. Instead, the improvement is more plausibly explained by the reduced skewness of mental health care costs that results from aggregating somatic and mental health care expenses into total expenses.

Comparison with previous research

These findings are consistent with those of van Asselt et al. (2014), who also reported a modest improvement in predictive accuracy when integrating short-term mental health care model with the somatic care model. In their study, the R^2 increased marginally from 0.223 to 0.224 – a larger improvement than observed in the present study where the R^2 increased from 0.3447 to 0.3449. The R^2 in their study is also more than 0.1 lower than the R^2 in this study, most likely because the somatic care model has been improved annually over the past ten years, leading to a higher R^2. However, unlike the current research, van Asselt et al. did not examine the effects of model integration on subgroup-level compensation, which is a key contribution of this research. In addition to the subgroup-level analysis, the present study extends the scope by having included long-term mental health care in the combined model. Even though the overall differences between the separate and combined models are smaller than those reported by van Asselt et al., the results still show a positive effect on predictive performance. This suggests that model integration remains beneficial, even when long-term mental health care is included – a dimension not previously addressed in literature.

Limitations and validity

Even though this study is considered a valid representation of real-world practice, several limitations must be acknowledged. First, individuals under the age of eighteen were excluded, as their mental health care costs are reimbursed under the Youth Act rather than the Health Insurance Act. In practice, however, health care costs of minors also need to be predicted for accurate risk equalisation. It would be inaccurate to assume that minors have no mental health care expenses for the model. Since the model aims to capture total health care costs, omitting mental health care expenses for individuals under eighteen results in an inconsistent dependent variable. This inconsistency would undermine the integrity of the combined model as it captures a different cost structure for minor than for adults. As somatic and mental health care costs may be correlated, this omission may also lead to biased or incomplete predictions. A potential solution would be to apply the previous somatic care model to individuals under eighteen, while using the combined model for the rest of the population. However, this approach would still require maintaining two models, potentially reducing the administrative and analytical advantages of adopting a unified model.

A second limitation of this study is the use of the OLS regression. OLS is generally less suited for predicting skewed distributions, and both somatic and – especially – mental health care expenditures are strongly right skewed. For this reason, in practice, somatic care costs have been predicted using constrained regression since 2024, while mental health care costs are modeled using quadratic programming (Cattel et al., 2024a). Although the models used in this study are not identical to the operational models applied in practice, they align with the broader methodological framework used in risk equalisation research. Specifically, the Dutch Ministry of VWS conducts all its research on the somatic care model using OLS regression, with constrained regression applied only in the final calculation stage at the end of the year (Smeets, 2024). In this context, the use of OLS in this study is consistent with standard research practice. Therefore, while there are valid arguments against using OLS for predicting highly skewed outcomes, its application in this study is justified. It ensures comparability with previous research and reflects the analytical approach most used in developing the Dutch risk equalisation system.

A third limitation of this study is that multiple classes had to be combined in order to perform the bootstrap procedure in Stata. Specifically, 35 original classes were merged into 15 new classes to address issues of collinearity that otherwise prevented model estimation in the bootstrap. The classes most affected were those that included age grouping, even the classes that were not in the age and sex risk-adjustor. This likely led to overly specific groupings with overlapping characteristics, thereby increasing the risk of collinearity. Notably, no strong correlations were found between the omitted classes and the other classes. The omission of classes in this analysis – unlike in the real-world model – may be due to the absence of constraints, which the actual model uses to address such issues. Nevertheless, combining classes had a small effect on the R^2. It changed by only 0.0002 for the combined model, as well as the separate models. Since the difference in R^2 between the combined and separate models was the same, the bootstrap results are still considered representative of the true difference in predictive performance between the two model types.

Since this study aims to reflect the real-world risk equalisation model, the R^2 of the somatic care model – including individuals under the age of eighteen – was compared to the R^2 reported in the Ministry of VWS's Overall Test from last year (Cattel et al., 2024b). Both models use the same specifications and data source, with the only difference being that the Ministry's model is based on the full national dataset (~17 million individuals), while this study used a 10% sample. However, the sample in this study was weighted to match the full dataset in terms of key demographic and risk characteristics, ensuring that the results are representative for the total population. In this study, the R^2 for the somatic care model with minors included was 0.320, compared to 0.323 in the real model. A direct comparison is valid here, as both models regress the same dependent variable on the same datasets using the same structure (van de Ven & van Kleef, 2024). The slightly higher R^2 in the Ministry's model can be attributed to the use of constraints, which were not applied in this research. Nevertheless, the R^2 values differ by only 0.003 indicating that the simplified version used in this study closely mirrors the predictive performance of the real model. This supports the conclusion that the model used in this research is a valid and useful approximation of the operational model used in practice.

Implications for policy

An important consideration is the complexity of the models, which has recently been explicitly included in the Ministry of VWS's model evaluation framework (Cluster risicoverevening, 2022). Combining the somatic and mental health care models may initially increase complexity, particularly during the transition phase. In the long-term, the model will contain all risk-adjustors from both existing models, which could suggest an increase in structural complexity. However, since the data sources remain unchanged, and the structure of how risk-adjustors are applied is also preserved, this complexity may be more limited than expected. Moreover, maintaining and updating a single model over time could reduce the overall workload, as only one model would need maintenance. This may help streamline development and prevent the risk of neglecting one of the separate models. Therefore, despite a possible short-term increase in complexity, the combined model could offer a more efficient and sustainable long-term solution.

Suggestions for future research

Based on the findings in this study, it is recommended that the Ministry of VWS considers combining the somatic and mental health care risk equalisation model into a single combined model. However, before implementation, several areas require further investigation. First, additional research should focus on the prediction of health care costs for individuals under the age of eighteen, who were excluded from this analysis. Accurate modelling for this population is essential for the risk equalisation for insurers. Second, further study is needed on the application of constraints within the combined model. This includes evaluating the influence of current constraints and exploring the potential impact of alternative constraints. Third, it would be valuable to investigate alternative estimation methods for the combined model, such as constrained regression or quadratic programming, to determine whether they offer improvements in predictive performance. Finally, it is recommended that the Ministry consult with experts in model complexity to carefully assess the trade-offs between increased model complexity and long-term efficiency, ensuring that any transition to a unified model is both feasible and sustainable.

Conclusion

In conclusion, this study demonstrates that combining the somatic and mental health care risk equalisation models can lead to modest yet meaningful improvements in predictive performance, particularly for chronically ill and those with high health care expenses. While statistical significance does not always translate into large gains in R^2, CPM or MAPE, the observed reduction in undercompensation for individuals with high mental health care costs indicates a potential for improved financial equity. This, in turn, could help reduce insurers' incentives for risk selection against these groups. Moreover, integrating the two models could lessen administrative burden and ensure more consistent model development. Although further research is needed – particularly regarding the prediction of health care expenses of minors, constraint application and alternative estimation methods – the results of this study suggest that transitioning to a combined model is a promising direction for future risk equalisation policy.

References

Agema, M. F. (2024, September 24). Herziening Zorgstelsel. Ministerie van VWS. Retrieved from https://zoek.officielebekendmakingen.nl/kst-29689-1266.html

Bundesministerium für Gesundheit. (n.d.). Risikostrukturausgleich (RSA). Retrieved April 14, 2025, from https://www.bundesgesundheitsministerium.de/risikostrukturausgleich.html

WOR 1222: Cattel, D., Eijkenaar, F., Panturu, A., van Kleef, R. C., & van Vliet, R. C. J. A. (2024a, September 30). Onderzoek risicoverevening 2025: Berekening Normbedragen. ESHPM.

WOR 1221: Cattel, D., Eijkenaar, F., Panturu, A., van Kleef, R. C., & van Vliet, R. C. J. A. (2024b, September 30). Onderzoek risicoverevening 2025: Overall Toets. ESHPM.

WOR 875: Cattel, D., Eijkenaar, F., van Kleef, R. C., van Vliet, R. C. J. A., & Withagen-Koster, A. A. (2017, August 8). Onderzoek risicoverevening 2018: Overall Toets. Instituut Beleid en Management Gezondheidszorg (iBMG).

WOR 1234: Cluster risicoverevening. (2024). Toetsingskader 2024 en verder. Ministerie van Volksgezondheid, Welzijn en Sport.

WOR 1130: Cluster risicoverevening. (2022, November). Toetsingskader 2022. Ministerie van Volksgezondheid, Welzijn en Sport.

Coralogix. (2023, August 12). A comprehensive Guide to Mean Absolute Percentage Error (MAPE). Retrieved from https://coralogix.com/ai-blog/a-comprehensive-guide-to-mean-absolute-percentage-error-mape/

Cumming, R. B., Knutson, D., Cameron, B.A., & Derrick, B. (2002) A Comparative Analysis of Claims Based Risk Assessment for Commercial Populations, Society of Actuaries, USA. https://www.researchgate.net/publication/228378875_A_Comparative_Analysis_of_Claims_Based_Risk_Assessment_for_Commercial_Populations

de Myttenaere, A., Golden, B., Le Grand, B., & Rossi, F. (2016, March 10). Mean Absolute Percentage Error for regression models. Neurocomputing. *Vol 192*, pp. 38-48. https://doi.org/10.1016/j.neucom.2015.12.114.

Douven, R. (2010). *Ex-postcorrectiemechanismen in de Zorgverzekeringswet: Hoe nu verder?*. (Report no. 212). Centraal Planbureau

Enthoven, A. C. & van de Ven, W. P. M. M. (2007, December 13). Going Dutch – Managed-Competition Health Insurance in the Netherlands. The New England Journal of Medicine. *Vol 357*(24), pp. 2421-2423. https://doi.org/10.1056/NEJMp078199

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021). An introduction to statistical learning: With applications in R (2nd ed.). Springer. https://www.statlearning.com

Kim, S. & Kim, H. (2016, July). A new metric of absolute percentage error for intermittent demand forecasts. International Journal of Forecasting. *Vol 32*(3), pp. 669-679. http://dx.doi.org/10.1016/j.ijforecast.2015.12.003

Linacre, J. M. (n.d.). PWEIGHT= person (case) weighting. Winsteps. Retrieved May 26, 2025, from https://www.winsteps.com/winman/pweight.htm

Ministerie van Volksgezondheid, Welzijn en Sport. (2019). Regeling risicoverevening 2020. Staatscourant van het Koninkrijk der Nederlanden (55238).

Nederlands Jeugdinstituut. (n.d.). Impact maken met jeugdbeleid. Retrieved February 4, 2025, from https://www.nji.nl/impact-maken-met-jeugdbeleid/jeugdwet

Nivel (2022). Nivel Zorgregistraties. 2022. NZR-00322.052.

Overbeek, L. (n.d.). Dit zijn we: Nivel Zorgregistraties Eerste Lijn. Nivel. Retrieved April 3, 2025, from https://www.nivel.nl/nl/panels-en-registraties/nivel-zorgregistraties-eerste-lijn/dit-zijn-we

PWC. (2023, November). *Innovatie in het risicovereveningsmodel: meer doelmatigheid en minder complexiteit* [Powerpoint slides]. Retrieved from https://www.pwc.nl/nl/marktsectoren/publieke-sector/documents/pwc-innovatie-in-het-risicovereveningsmodel-meer-doelmatigheid-en-minder-complexiteit.pdf

Smeets, J. (2024, August 19). Memo Uitgangsmodel bij partiële onderzoeken. Ministerie van Volksgezondheid, Welzijn en Sport.

Stat20. (n.d.). Bootstrapping: Another Approach to Confidence Intervals. Stat20 UC Berkeley. Retrieved May 28, 2025, from https://www.stat20.org/4-generalization/03-bootstrapping/notes

StataCorp LLC. (2019). Stata Base Reference Manual Release 16. Stata Press

WOR 726: van Asselt, M., Ellwanger, N., Faber, D., & Everhardt, T. (2015, March 10). Doorontwikkeling van het risicovereveningsmodel voor de gGGZ en de lGGZ. Ape.

van de Ven, W. P. M. M., Beck, K., Buchner, F., Schokkaert, E., Schut, F. T., Shmueli, A., & Wasem, J. (2024). Preconditions for efficiency and affordability in competitive healthcare markets: Are they fulfilled in Belgium, Germany, Israel, the Netherlands and Switzerland? Health Policy, 143, 105099. https://doi.org/10.1016/j.healthpol.2024.105099

van de Ven, W., Hamstra, G., van Kleef, R., Reuser, M., & Stam, P. (2022, March 29). The goal of risk equalization in regulated competitive health insurance markets. *Eur J Health Econ 24*, pp. 111–123 https://doi.org/10.1007/s10198-022-01457-7

van de Ven, W. P. M. M. & van Kleef, R. C. (2024). Why do we use R^2 so often in risk equalization research? EsCHER Working Paper Series No. 2024001, Erasmus University Rotterdam. https://www.eur.nl/en/research/research-groups-initiatives/escher/research/working-papers

Vanhommerig JW, Verheij RA, Hek K, et al. Data Resource Profile: Nivel Primary Care Database (Nivel-PCD), The Netherlands. International Journal of Epidemiology, 2025, 54(2), dyaf017 https://doi.org/10.1093/ije/dyaf017

van Kleef, R. C., Eijkenaar, F., van Vliet, R. C. J. A., & van de Ven, W. P. M. M. (2018). Chapter 14 – Health Plan Payment in the Netherlands. In Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets, pp. 397-429. *Elsevier*. https://doi.org/10.1016/B978-0-12-811325-7.00014-2.

van Kleef, R. C., Reuser, M., McGuire, T. G., Armstrong, J., Beck, K., Brammli-Greenberg, S., Ellis, R. P., Paolucci, F., Schokkaert, E., & Wasem, J. (2024). Scope and Incentives for Risk Selection in Health Insurance Markets With Regulated Competition: A Conceptual Framework and International Comparison. *Medical Care Research and Review, 81*(3), pp. 175-194. https://doi.org/10.1177/10775587231222584

van Kleef, R. C., van de Ven, W. P. M. M., & van Vliet, R. C. J. A. (2013a). Risk selection in a regulated health insurance market: a review of the concept, possibilities and effects. *Expert Review of Pharmacoeconomics & Outcomes Research, 13*(3), pp. 743-752. https://doi.org/10.1586/14737167.2013.841546

van Kleef, R. C., van de Ven, W. P. M. M., & van Vliet, R. C. J. A. (2013b). Risk equalization in The Netherlands: an empirical evaluation. *Expert Review of Pharmacoeconomics & Outcomes Research, 13*(6), pp. 829-839. https://doi.org/10.1586/14737167.2013.842127

Vektis. (n.d.). Tijdlijn 15 jaar Zorgverzekeringswet. Retrieved February 4, 2025, from https://www.vektis.nl/intelligence/publicaties/tijdlijn-15-jaar-zorgverzekeringswet

Zheng, C., Yin, J., Wu, L. Hu, Z., Zhang, Y., Cao, L., & Qu, Y. (2024). Association between depression and diabetes among American adults using NHANES data from 2005 to 2020. *Scientific Reports, 14*, 27735. https://doi.org/10.1038/s41598-024-78345-y

Appendix

Appendix 1:

Omitted classes per model

Combined model	Somatic care model	Mental health care model	Class definition
LG 18	LG 18	LG 18	Men; age 75-79
SES 6	SES 6	SES 6	Low socioeconomic status; age 70+
AVI 6	AVI 6	AVI 6	IVA; age 65-69
AVI 15	AVI 15	AVI 15	Income support; age 35-44
AVI 16	AVI 16	AVI 16	Income support; age 45-54
AVI 17	AVI 17	AVI 17	Income support; age 55-64
PPA 9	PPA 9	PPA 9	In a long-term care facility with treatment, newly admitted; age 80+
PPA 11	PPA 11	PPA 11	In a long-term care facility without treatment or extra-mural long-term care, newly admitted; age 70-79
Mental FKG2			Addiction excluding nicotine addiction

Combined classes	New classes
AVI 2 & 3	avi2_3
AVI 4 & 5 & 6	avi4_5_6
AVI 9 & 10	avi9_10
AVI 11 & 12	avi11_12
AVI 15 & 16	avi15_16
AVI 17 & 18	avi17_18
AVI 29 & 30	avi29_30
LG 17 & 18 & 19	lg17_18_19
LG 20 & 21	lg20_21
LG 37 & 38 & 39	lg37_38_39
LG 40 & 41 & 42	lg40_41_42
PPA 2 & 3	ppa2_3
PPA 7 & 8 & 9	ppa7_8_9
PPA 11 & 12	ppa11_12
SES 5 & 6	ses5_6