‘To What Extent Can Prior use of Youth Care Improve the Dutch Risk-equalisation Model under the Health Insurance Act (Zvw)?

Erasmus University of Rotterdam

Master Thesis Research

Version: 4 November 2026. A previous version (published between 29 September 2026 and 4 November 2026) contained an error in the ordering of the labels for ‘type of youth care’ in all relevant tables. This error has been corrected in the current version.

Erasmus University of Rotterdam – Faculty of Health Policy and Management
Student name: Maryam Diallo
Master name: Msc. Health Economics, Policy & Law
Student number: 785294
Due date: 8 June 2026 Location: Rotterdam
Supervisor name: Richard van Kleef
Mail : 785294md@student.eur.nl
Word count : 14,717

Acknowledgement:

I am grateful to the Dutch Ministry of Health, Welfare and Sports and the Association of Health Insurers for access to (anonymized) claims data. Next, I also want to show my gratitude to my supervisor, Dr. Richard Van Kleef, who has guided me with knowledge and know-how of the subject matter, as well as with a clear structure and methodology. Lastly, I also want to thank God, my family and my friends, who have supported me throughout these months of writing my thesis.

AI statement :

AI has been used in this thesis as tool for the coding parts on Rstudio, for explaining terminology, abbreviations and concepts, and helping with setting up tables and formatting. Grammarly has been used for checking spelling.

Abstract

Objectives:

This study aims to answer the research question; 'To what extent can prior use of youth care improve the risk equalization model under the Health insurance act (Zvw)?'. This is done by researching the following sub-questions: (1) 'What are the characteristics of users of youth care?' (2) 'Does prior youth care predict total Zvw costs?'; (3) ''Does prior youth care predict residual Zvw costs?'; (4) 'Does youth care meet the qualitative criteria for a risk adjuster?'.

Background:

The Dutch health insurance market uses regulated competition, where insurers are compensated by the risk equalization model for the predictable differences in healthcare costs among enrollees. With over 200 existing risk adjusters, existing gaps of undercompensated enrollees remain not covered.

Methods:

This study employs microdata from the CBS, covering 4.9 million insured individuals under 25 in the Netherlands. This study employs three datasets: Zvw somatic costs, Zvw GGZ costs (mental healthcare), and youth care historical trajectories from 2020, 2021 and 2022. The analytical approach includes descriptive analysis, OLS regressions, residual cost computation and a qualitative criteria assessment.

Results:

The empirical results show that youth care users are systematically disadvantaged compared to non-youth care users, with lower socioeconomic status and higher morbidity. Second, youth care shows statistically significant, but individually weak predictive power on total Zvw costs. Third, the subgroups of youth care types are all undercompensated, meaning insurers face predictable losses on the enrollees. Lastly, youth care partially meets the requirements for inclusion, indicating that incentives are set right, but collection of data and its quality are hurdles that push back its integration.

Conclusions:

Prior use of youth care can potentially improve the risk equalization model. In order to integrate youth care as risk adjuster, it is recommended to ensure a uniform and stable registration of information on youth care data.

Table of Contents

1. Introduction 6

2. Theoretical Framework 8

2.1 Regulated competition in the Dutch Health Insurance Market 8

2.2 The Role and Relevance of risk equalization 9

2.3 Youth Care 12

2.4 Criteria for risk adjustors 13

3. Methodology 15

3.1 Data sources 16

3.2 Data execution 17

3.2.1 Characteristics of Youth Care Users 18

3.2.2 Predictive power for Total Zvw Costs 18

3.2.3 Predictive Power for Residual Zvw Costs 19

3.2.4. Qualitative Assessment of Risk Adjuster Criteria 19

4. Results 20

4.1 Characteristics of youth care users 21

4.2 Predictive power for total zvw costs 26

4.3 Predictive power of residual Zvw costs 30

4.4 Qualitative assessment 34

5. Discussion and Conclusion 36

5.1 Summary of main key findings 36

5.2 Strengths, Limitations and Research recommendations 38

5.3 Policy implications 39

5.4 Conclusion *..... 40*

6. References 41

7. Appendix 44

1. Introduction

The health insurance market in the Netherlands is characterized by a competitive landscape with competition between insurers and care providers. Its current landscape is organized according to the principles of regulated competition, rooted in the the implementation of the health insurance act (Zvw) put in place in 2006. This act mainly stemmed from the increasing need for better accessibility, higher quality and affordable prices in health care (Enthoven & Van De Ven, 2007). The act obliges all citizens to take out health insurance. While allowing enrolees with a free choice of choosing health care insurers on a contractual and annual basis, insurers became obliged to set a nominal premium for all enrolees, irrespective of age, gender or health status (Borkent, K., et al., 2016).

The tightly regulated health insurance market in the Netherlands is thanks to decades of regulatory improvement on both demand- and supply-side interventions. First, what sets apart the new era of the Dutch health insurance market is the focus on the 'mobility' of the insured, which has been proven to be central for improving the competitive landscape for insurers (Rooijen et al., 2011). This is because the satisfaction level of enrollees depends on the premium and quality of care provided by the insurer. Resultantly, if an insured is dissatisfied with their current insurer, this poses an opportunity for the next best insurer to provide a service better than its competitor. On the supply-side, the Dutch regulatory environment has also introduced various mechanisms and tools that benefit and smoothen the work of insurers. The risk-equalization scheme, who had been introduced in the 90s serves as crucial tool for compensating health insurers for predictable variation in their enrolled individual medical expenses (Van Kleef et al., 2013).

Over time, the risk-equalization model has adopted a wide variety of variables, ranging from demographic to morbidity-based characteristics deemed as important for estimating the compensation to the insurers (Van Kleef et al., 2013). By now, the model counts over >200 risk classes, each classified according to three sub-models, namely somatic health care, mental health care, and out-of-pocket payments (Van Kleef et al., 2018). Certain groups, however, remain either under- or overcompensated, or unidentified in the risk-equalization scheme. This means that there is still room for improvement and identification for new variables. Hence, to improve the accuracy and coverage of the risk equalisation scheme, this research will study whether prior use of youth care can improve the risk equalization estimates.

Youth care, which lies its foundation in the Child and Youth Act (2015), provides the necessary care to children up to the age of 18 who suffer from psychological, psychosocial, behavioural and intellectual disability (Netherlands youth Institute, 2025). Youth care focuses on providing supervision, with services offered ranging from rehabilitation, to ambulatory care, foster care and more (Ministerie van VWS, 2025). Understanding the potential value of youth care on the risk equalisation model is

relevant for the following reason. First, if a parent is expected to take out long-term child protection or psychological care for their child before being reaching the age of 18, the child might incur above-average costs for the insurer than a typical 18-year old. This is because when the patient reaches the age of 18, the individuals filed information on support and costs get transferred under different care laws, under which the Health Insurance act (Zvw), Long-term care act (Wlz) and Social Support act (Wmo). Since youth care does not exist as a risk-adjuster, insurers are not yet compensated accordingly for this transition. Hence, previous use of youth care in the previous year (t-1) might have a correlation with the same usage under Zvw. Therefore, we expect that previous use of youth care in the previous year to have a predictive influence on costs under Zvw.

The aim of this research is to give an answer to a set of questions that will be researched quantitatively and qualitatively, to assess the appropriateness of youth care as risk-adjuster. Therefore, the main research questions of this thesis is: 'To what extent can prior use of youth care be used as a risk adjuster to improve the Dutch risk-equalisation model under the Health Insurance Act (Zvw)?'. In addition, this research will employ a set of sub-questions, where the first two provide an explanatory analysis to get insight on the relationship between previous utilization of youth care in the prior year (standardized) and costs under health insurance (Zvw):

(1) What are the characteristics of people with prior use of youth care?
(2) To what extent is prior use of youth care predictive of total health care spending under the health insurance act (Zvw)?
(3) To what extent is prior use of youth care predictive of residual health care spending under the health insurance Act, i.e. expenditures remaining after risk equalisation?
(4) To what extent does prior use of youth care fulfil the criteria for risk adjusters?

The outline of this thesis will go as follow: first, the thesis will discuss the relevant theoretical frameworks on regulated competition, risk-equalization and youth care. Then, by using the relevant data on youth care and health insurance (under Zvw), the empirical tests will be done to observe any significant results on the predictive power of youth care. Lastly, after analyzing the results, this research will discuss the summary of key findings, the strengths and limitations, as well as recommendations for future research and policy implications, followed by the conclusion.

2. Theoretical Framework

The theories used in this research paper are grounded in the foundations of regulated competition in health insurance. The aim of this research is to understand whether prior youth care use has a predictive value on actual Health Insurance care costs (Zvw). Therefore, this section serves as relevant framework for guiding readers on the relevant research topics at hand.

2.1 Regulated competition in the Dutch Health Insurance Market

2.1.1 Competition

Competition between health insurers within the Dutch health insurance system has been a crucial component for driving efficiency and quality of service in the market. The conditions in which health insurers are allowed to compete on have been tightened through the implementation of the health insurance Act mandated in 2006. The act merged both statutory and other private health insurance schemes under one universal scheme, allowing for a mandatory but free choice of insurer (Jeurissen & Maarse, 2021). The main mechanisms introduced in the reform included the responsibility for insurers to act as purchaser of health services on behalf of the consumers. By giving insurers the responsibility to manage the quality of care of the provider network and out-of-pocket payments, the policy reforms put in place allow to improve equal efficiency in delivery of care (Van Kleef & Van Vliet, 2025).

To stimulate competition in the health insurance market, insurers are allowed to compete on both price and quality. First, they are allowed to compete on price by setting competitive premiums differently from each other. However, in some cases insurers can put their own interest instead of acting in the best interest of their enrollees. For instance, in certain areas of health care like dentistry, patients had complained about the increase of treatment prices, adversely influencing the use of certain types of care (Trescher et al., 2019; NZa & Langejan, 2012). Therefore, the health insurance act also aims at ensuring quality of care, to counter for misaligned incentives that could occur with price competition. To ensure quality of care, insurers try to negotiate on behalf of their enrollees, as well as include agreements on ensuring good quality of care. By actively purchasing and negotiating with third-parties, insurers are also allowed to apply selective contracting with providers of care, allowed under the Health Insurance Act (2006).

2.1.2 Regulation

The establishment of the Health Insurance Act (2006) obliges everyone that qualifies as a legal resident or employee in the Netherlands to purchase individual private health insurance. On the supply-side, private health insurance companies are obliged to accept each applicant during the open enrollment period, regardless of the enrollee's previous pre-existing conditions (Van De Ven & Schut, 2008). On

the demand-side, enrollees have the option to switch to another insurer once per year, as each contract period has a duration of at most one year. Contracts must include precise entitlements between the applicant and the insurer, including the list of contracted providers, the covered pharmaceutical and other entitlements toward the individual (Van De Ven & Schut, 2008). The act also explicitly prohibits premium differentiation ban, obliging that premiums must be the same for each enrollee, independent from age, gender or health status.

The premiums for each enrollee are written at a community-rated level per product on the contract, which are either paid through income-related contributions (tax contributions), or the nominal premium paid directly to the insurer. These payments occur through taxable income for employees or are also funded by the government to Dutch households through income-related subsidies, referred to as healthcare allowance (in Dutch: zorgtoeslag) (Van De Ven & Schut, 2008). Children under the age of eighteen are taken under the responsibility of the government, under which a contribution is made on their behalf to the Risk-Equalization Fund (REF).

2.1.3 Risk selection

Regulated competition combats cream-skimming behaviour, which involves the active selection of consumers who are often low-risk and highly profitable (Barros, 2003; Van Kleef et al., 2024). An example of cream-skimming is product differentiation, where insurers design different insurance plans to attract more healthy individuals instead of attracting the chronically ill (Van Kleef et al., 2024). The consequences of cream-skimming defy the core principles of the Health Insurance Act (2006) through defying solidarity within the risk pool of an insurer, as well as decreasing equity of access for high-risk enrollees. These factors reflect the existing competitive landscapes that demand further control of the Dutch health insurance market. To combat those inefficiencies, Dutch authorities adopted the risk equalization scheme as the core mechanism to prevent cream-skimming under the Health Insurance Act. This passage reveals one crucial aspect of why it is important to research the mechanisms and possible risk adjusters that could possibly improve the risk equalization scheme, as well as the overall competitive health insurance landscape in the Netherlands.

2.2 The Role and Relevance of risk equalization

2.2.1 Goals of Risk-equalization

The implementation of the Risk equalization scheme has been a major contributor for regulating the competitive landscape of the health insurance market. Its implentation in 1993 was to achieve multiple goals that would benefit both insurers and consumers of health insurance (Van De Ven et al., 2022). First, the goals of risk equalization are to achieve an ‘equalising effect’ for insurers,which refers

to the situation when each insurer's expected portfolio in year t is independent from the risk composition of his portfolio in year t. At that stage, the insurers reach a level playing field, where incentives for risk selection are also decreased, and insurers have almost the exact same portfolio, forcing hem to compete on factors like quality of care (Van de Ven et al., 2022). This situation is crucial to achieve efficient competitive, because in such way, insurers have an identical expected financial result, and reduce the incentives for risk selection.

2.2.2 How Risk Equalization works

The risk equalisation framework happens through multiple steps. First, researchers refer to 'ex-ante risk equalisation', meaning that the compensation to insurers is estimated with predicted costs, before the actual costs are known (Van Kleef et al., 2018). So the risk equalization model looks at the characteristics of the enrolled, which vary across factors like age, gender or PCGs and DCGs etc. Then, following the estimation, or after risk equalisation, the expected cost per insurance product is estimated and redistributed to insurers. Based off the insurer's portfolio, an adequate compensation is given the higher the risk composition of the pool is. Similarily, insurer's with a portfolio consisting of low-risk enrollee's will generally speaking receive a lower payment. The main objective is to neutralise incentives for cream skimming, as well as stimulate a level playing field for insurers.

2.2.3 Existing Risk adjusters

Table 1 presents an overview of each of the existing risk adjusters for the ex-ante risk-equalisation scheme in 2026. Each characteristic is classified across the different sub-models,which helps at having a more accurate prediction of costs from different parts of healthcare. First, the somatic model focuses on physical health care, which includes hospital visits, physiotherapy etc. Next, the mental healthcare model mainly focuses on mental healthcare cost and are mostly characterized by patients with high needs for healthcare, hence incurring larger costs for insurers (Van Kleef et al., 2018). This could be factors like Pharmacy-Based Cost groups (PCGs), Diagnosis-Based Cost groups (DCGs), Physiotherapy Diagnosis Groups (PDG) or Multiple-year nursing & care costs (MYNC).

The third sub-model is the out-of-pocket payments, where the usage of care of a person within a year is tracked. More specifically, by focusing on the threshold of the deductible which is €385 in the Netherlands, the risk adjusters are limited to Age and Gender, Source of income, multiple year high costs, and seasonal worker. Each of these sub-models receives funding through different streams, as well as focuses on different networks of providers (Van Kleef et al., 2018).

Table 1. Existing Risk adjusters for Somatic care, GGZ and deductible payments (for the risk-equalisation scheme in 2026)

Type of Risk adjuster	Number of risk classes		
	Somatic model	GGZ-model	Out-of-pocket payments
Age and gender (Age/Gender)	42	30	30
Pharmacy Costs Group (PCG)	50 + 1*	9 + 1*	---
Diagnostic Costs Group (DCG)	26 + 1*	16 + 1 *	---
Source of income* Age	36	29	29
Regional zip-code clusters (ZIP)	10	10	10
Socioeconomic status* Age	12	8	---
Persons per address*age	19	18	---
Multiple year high/low costs (MCHGs)	8+1*	7+1*	1+1*
Physiotherapy diagnosis groups (PDG)	4+1*	---	---
Multiple-year nursing & care costs (MNC)	9+1*	---	---
Historical somatic morbidity	1+1*	---	---
Deliveries /pregnancies	3+1*	---	---
Seasonal worker	3	3	3
Historical medical equipment costs	6+1*	---	---
Missing costs history	4+2*	---	---
Total number of risk classes	**243**	**133**	**74**

Source: Van Kleef, R., R.C.J.A. Van Vliet, & ESHPM, Rotterdam: Erasmus Universiteit. (2026). Verevenende Werking & Power. *WPR-analyse Risicoverevening 2026:* https://www.eur.nl/eshpm/media/2026-05-eindrapport-wpr-analyse-rv26-v30apr2026-1

*+1 reflects the residual category, i.e. the category for the insured without a score on the relevant morbidity criterion. For example, the PCGs refer to the enrollees who have not been classified into any of the 49PCG categories (Van Kleef, R. et al., 2026)

2.3 Youth Care

In 2022, 467,000 people under 23 received youth care, reflecting about 10.5% of all young people within that age group (Ministerie van VWS, 2025). Youth care is a service under the Child and Youth Act (Jeugdwet) introduced in 2015, obliging municipalities to organise and finance care for young people in need (Netherlands Youth Institute, 2019).

Table 2 writes out the existing youth care types that officially have been recorded within the government system. In 2022, 'outpatient care at the provider's location' was the most common form of youth care, representing 70% of all youth care recipients in the Netherlands (Bakker, 2025).

Table 2. Type of youth care under CBS data

	Dutch description	English description
Hulp_01	Jeugdhulp zonder verblijf, uitgevoerd door wijk- of buurtteam	Youth care without residential stay, provided by a district or neighbourhood team
Hulp_02	Jeugdhulp zonder verblijf, ambulant, op locatie van de aanbieder	Youth care without residential stay, outpatient care at the provider's location
Hulp_03	Jeugdhulp zonder verblijf, daghulp op locatie van de aanbieder	Youth care without residential stay, day care at the provider's location
Hulp_04	Jeugdhulp zonder verblijf, jeugdhulp in netwerk van de jeugdige	Youth care without residential stay, care provided within the young person's social network
Hulp_11	Jeugdhulp met verblijf, pleegzorg	Youth care with residential stay, foster care
Hulp_12	Jeugdhulp met verblijf, gezinsgericht	Youth care with residential stay, family-based care
Hulp_13	Jeugdhulp met verblijf, gesloten plaatsing	Youth care with residential stay, secure residential placement
Hulp_14	Jeugdhulp met verblijf, anders dan bovenstaand	Youth care with residential stay, other than the categories listed above

Source: CBS (2025): https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/maatwerk-microdatabestanden/maatwerk-microdatabestanden/jeugdstelsel/jgdhulpbus-trajecten-jeugdhulp-aan-jeugdigen

Jeugd-GGZ, which is youth mental health care, provides younger people with specialist psychological and psychiatric support to children and young people under 18 with diagnosed needs. Since the implementation of the Youth care act in 2015, jeugd-ggz has also become under the responsibility of municipalities to care for (Ministerie van VWS, 2025). Interestingly, the number of young people using youth care has increased by 23% between 2015 and 2022, with mental health reflecting the main driver for this increase in usage (Kroneman & Beerman, 2024).

2.3.1 Youth Care Beyond 18

Upon turning 18, individuals are no longer covered by municipal youth care budgets. Under normal circumstances, they now must rely on the basic health insurance package under the Zvw. However,

debates on the extension of youth care have been made have lead to some forms of youth care assistance like foster care has been allowed up to the age of 21. This is because there have been cases in which users of mental health care face elevated risks of mental health problems, substance use disorders, chronic conditions in adult life, each associated with higher health care utilisation (Netherlands Youth Institute, 2019). Over time, the Dutch government has recognised this higher vulnerability to the age group of 16 to 27 age group in multiple policy documents, giving rise to the concept of 'integrated youth care transition' and the statutory right to extended youth care up to age 23 (Ministerie van Volksgezondheid, Welzijn en Sport, 2025).

2.4 Criteria for risk adjustors

This section outlines the principal criteria for evaluating (potential) risk adjustors drawn from Van de Ven & Ellis (2000), and the criteria from the Dutch WOR assessment framework. More specifically, a risk adjuster must satisfy a broader set of normative and practical criteria that reflect the dual objectives of the risk equalisation system: eliminating incentives for risk selection while preserving insurer incentives for cost efficiency. The decision to include a new variable as a risk adjuster in the Dutch risk equalisation system is not made on the basis of predictive power alone.

2.4.1 Quantitative measures of model performance

To evaluate the fit of a risk adjuster, various quantitative measures exist. First, the R^2 is a commonly used measure within in risk equalisation research. It is defined by 'the proportion of the total variance in individual healthcare expenses that can be predicted by risk adjusters'[1] (Van Kleef & Van De Ven, 2024). Formula 1 represents a ratio of the sum of squared differences between actual and predicted costs and ranges between 0 and 1, where a value close to 0 indicates the model explains little variation, and a value near 1 explains almost all of it.

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

However, previous research has highlighted certain limitations for using the R-square as primary measure. One limitation of using R-square is that the model automatically inflates the R^2, and does not have any specific benchmark for interpreting its value, as the value will always be lower than 1 (Van Kleef & Van de Ven, 2024).

For this reason, research makes use of other quantitative measures as complementary measure for interpreting the fit of a risk adjuster, such as the mean residuals at the subgroup level. This measure

[1] This definition is based on research within health insurance. The R^2 measure can take slightly different definition when applied to its specific statistical context.

is encouraged by research because of its ability to measure the degree to which the system actually achieves its primary objective of eliminating predictable profits and losses on identifiable groups of enrollees (Van Kleef & Van de Ven, 2024). Formula 2 represents the mean financial result, which is the standard meaasure for assessing whether a model eliminates preditable profits and losses.

$$(2) \quad \text{Mean financial result}_g = \frac{\sum_{i=1}^{g}(\hat{Y}_i - Y_i)}{n_g}$$

Further, research encourages the use of this measure because it considers the mean residual effect at a subgroup level, which the R^2 ignores (Van de Ven & Van Kleef, 2024). The interpretation follows that a positive outcome reflects overcompensation for the insurer, hence translating into predictable profits. On the contrary, a negative outcome reflects undercompensation, meaning that the group generates a predictable loss for the insurer. In addition, if the mean residual equals 0, the insurer incurs neither a loss nor a gain from the subgroup. In this study, the residual cost for each individual, defined as actual Zvw cost minus predicted Zvw cost will be used as the outcome variable in Step 3 of the analysis to assess whether prior youth care use is associated with systematic under- or overcompensation under the current model. In the empirical analysis, the costs are calculated by the actual minus the predicted cost. For example, residual cost (for somatic care) = actual cost (for somatic care) – predicted cost (for somatic care). This is the inverse of the mean financial result formula from section 2.4.1. The results are also meant to be interpreted the other way around, with a positive residual indicating undercompensation, whilst a negative residual indivates overcompensation.

2.4.2 Appropriateness of incentives

Van de Ven & Ellis (2000) propose a set of useful criteria to compare the different risk adjustment models. For instance, cost differences should reflect N-type factors (need-based differences deemed unavoidable) or S-type factors (efficiency-driven differences, where risk is set on insurers for providing efficient care). A risk adjuster shapes insurer behaviour in two ways simultaneously: it reduces incentives for risk selection (a benefit), and it potentially weakens incentives for cost efficiency (a cost).

2.4.3 Fairness

Fairness refers to the specific decision of which risk factors should be labeled as an S-type (eg. Lifestyle choices) or N-type factors (eg. age, health status), influenced by what society deems to prioritise. For instance, costs that are due to lifestyle factors, like smoking, might be non-subsidised, because these expenses are brought by the individual (Van de Ven & Ellis., 2000). On the contrary, people might argue that these costs should be compensated for, as the individual consequence of becoming ill could be greater, and receiving appropriate medical treatment is deemed as a priority. In

the context of youth care, the fairness criterion asks whether society should pay for the cost differences of youth care users. Central to the criteria lies the question of whether the costs incured by youth care users occur from a place of need (N-type factor), which should then be compensated by the risk-equalisation model (Van de Ven & Ellis, 2000).

2.4.4 Complexity

The WOR framework addresses feasibility as an overarching criterion that covers multiple aspects: simplicity & transparency, stability and availability & relaibility. First, the WOR framework defines stability as the ability of a risk adjuster to not fluctuate too much from year to year in terms of compensation (Van Kleef et al., 2026). Next, Simplicity & transparency refer to whether the risk adjuster has data that is independent from the influence of insurers or providers, as well as whether its cost coefficients are deemed as sufficient to interpret by stakeholders for its possible inclusion (Van Kleef et al., 2023).

2.4.5 Validity

The validity criterion the criterion aims to measure whether the potential variable measures the actual health need, without distortion (Van Kleef et al., 2026). In addition, the potential risk adjuster must be recorded in a reliable way, without measurement errors across the data of the enrollees (Van Kleef et al., 2023). Lastly, the WOR framework refers to validity as the ability of the variable to not be manipulable by insurers or providers.

2.4.6 Feasibility

The feasibility refers to whether the variable can be implemented in the RE-model (Van de Ven & Ellis, 2000). This criterion aims to look at whether the data is available on a recuring and fixed basis, and whether the data can be obtained without complexity. Essentially, this criterion aims to look at whether the variable can be collected and operationalised within the existing RE framework.

3. Methodology

The methodology section serves as descriptive overview of the data and analytical approach used in this research. Section 3.1 introduces the data sources used for the quantitative analysis. Then, section 3.2 explains how the subquestion of the underlying research will be answered analytically.

3.1 Data sources

For this research, three national data files are used, with each one containing data on the 2026 somatic care and out-of-pocket risk equalisation dataset, one on the 2026 GGZ risk equalisation dataset, and one on youth care use files for 2020, 2021 and 2022. Each file contains the actual costs from 2023 as reference year, as well as the predicted costs from the 2026 risk equalisation model. An overview of each of the variables can be viewed in the Appendix table 1. Next, each file includes 8 care type dummies, containing each type of youth care that can be taken, as showcased in table 2, and refered to as 'hulp_type' in the dataset. However, in each table of this research the full name has been used.

The categories of variables of the datasets are divided according to demographic variables, somatic and psychiatric morbidity indicators, as well as socioeconomic and household variables. A complete list of each of the variables are displayed in Table 2. Lastly, the quantitative parts of the analysis are executed on Rstudio, ran through the CBS microdata environment (Central Bureau for Statistics).

Table 2 represents the descriptive output results, presented with the mean and standard deviation of each of the risk adjusters available in the dataset. The results of the descriptives show that the mean age of the population is 12.6 years, confirming that the population of research is in fact young, or under 21. Other mean results show that across the three years, the proportion of youth care use has been stable, with 9.4% in 2022, 9.2% in 2021, and 8.4% in 2020.

Table 2. *Descriptive statistics of youth care users (reference year 2023)*

Variables	**Mean**	**SD**
Age	12.6053	7.1374
Morbidity-based factors		
'At least one mPCG'	0.1112	0.3100
'At least one pPCG group in prior year'	0.0108	0.1018
'At least one pDCG group in prior year'	0.0265	0.1582
'At least one pMYHC group in prior year'	0.0364	0.1845
'At least one PCG'	0.0370	0.1862
'At least one DCG'	0.0257	0.1560
'At least one MYHC'	0.3344	0.4651
'At least one PDG'	0.0420	0.1978
'At least one MYNC'	0.0032	0.4651
Historical high costs	0.0155	0.1219
Socioeconomics status (SES)		
Very low	0.2329	0.4168
Low	0.1886	0.3857
Middle	0.2785	0.4420
High	0.3000	0.4518
Persons per address (PPA)		
Reference	0.6864	0.4574

One person	0.0158	0.1229
Other household type	0.2935	0.4490
Wlz institutional care, below threshold	0.0013	0.0352
Wlz non-institutional care, below threshold	0.0026	0.0506
Wlz institutional care, above threshold	0.0001	0.0090
Wlz non-institutional care, above threshold	0.0002	0.0150
Cost specifics (€)		
Total Zvw Healthcare costs	1577	10298
Actual somatic care costs	1443	9876
Actual mental health care costs	187	2817
Actual out-of-pocket costs	54	122
Residual Zvw costs	-82	9172
Youth care specifics		
Youth care user in any of the prior years (2020 – 2022)	0.1429	0.3451
Youth care types		
Youth care without residential stay, provided by a district or neighbourhood team	0.0135	0.1136
Youth care without residential stay, outpatient care at the provider's location	0.0685	0.2490
Youth care without residential stay, day care at the provider's location	0.0072	0.0836
Youth care without residential stay, care provided within the young person's social network	0.0209	0.1411
Youth care with residential stay, foster care	0.0042	0.0638
Youth care with residential stay, family-based care	0.0014	0.0363
Youth care with residential stay, secure residential placement	0.0004	0.0192
Youth care with residential stay, other than the categories listed above	0.0039	0.0616
Youth care dummies (t-1, t-2, t-3)		
Any youth care use in 2022 (t-1)	0.0941	0.2878
Any youth care use in 2021 (t-2)	0.0918	0.2847
Any youth care use in 2020 (t-3)	0.0839	0.2734

Note. Reference year 2022 - 2023. N = 4,877,025 except for the variable age (n = 4,877,007). The population covered individuals under 25. Binary variables have been reported as proportions (eg. 0.14 = 14%), while continuous variables are reported as means in original units (€ for costs, years for age). Costs have been rounded to full euro's. 'Wlz (non-)institutional care, below/above threshold' refers to whether some registered in someone's household is receiving long-term care (Wlz). Below or above refers to the number of people receiving care at the address. 'At least one X' refers to the binary 0/1 dummies in the dataset: a person has at least one classification in that specific group. Variable names: mPCG = Multiple-year high costs, pharmacy-based cost group ; pPCG = Pharmacy-based cost group in prior year ; pDCG = Diagnosis-based cost group in prior year ; pMYHC = Multiple-year high cost group ; PDG = physiotherapy diagnosis group ; MYNC = Multiple-year nursing and personal care costs group.

The mean results for the overall morbidity of the population is low, with 3.7% having at least 1 PCG, and 2.6% at least one DCG. This is in line with the research subject focusing on younger people, thus showing a low morbidity rate is to be expected. Next, the mean total Zvw costs are €1,577, of which somatic costs incurs the largest share amongst the submodels. The mean residual Zvw cost, which is -€82.1, indicates that the full sample under 25 is slightly overcompensated on average.

3.2 Data execution

To give a concise answer to the research questions, the analysis is structured around the four sub-questions part of the research question. The analytical approach consists of multiple steps across

the questions: descriptive profiling, OLS regressions, mean residuals computation and qualitative criteria assessment. This section provides one subsection for the approach per sub-question.

For purposes of the analysis, all costs in the file have been converted to an annual basis. For instance, for an insured person with actual costs of 2,000 euros and an enrollment period of 6 months (=0.5 years), the costs have been increased to 2,000/0.5 = 4,000 euros. In short: the file therefore contains the costs per insured year. A weight ("gew") has also been included in the file. This variable represents the fraction of the year that an insured person was enrolled with an insurer in 2023. Gew<1 occurs, for example, for people who were born, died, or migrated in 2023.

3.2.1 Characteristics of Youth Care Users

The first research question examines what the characteristics are of people with prior use of youth care. To answer this question, the first analysis researches the characteristics of youth care users compared to non-users in year t-1, or youth care use in the prior year. To understand the characteristics of youth care users, a second analysis is done focusing on youth care users by type of care received.

3.2.2 Predictive power for Total Zvw Costs

The second research question looks at whether prior use of youth care has predictive power on total health care spending for total Zvw costs. To answer this question, the first part of the analysis is a weighted mean cost comparison, separating between the three submodels (total Zvw, somatic and GGZ). Next, the analysis includes two (Ordinary Least Squares) regression analysis, one base model focusing on costs from the year prior, and one extended model accounting for the costs in year t-2 and t-3. In this analysis, the primary summary measure of the predictive power for youth care use is measured using R^2. This captures the proportion of variance in Zvw costs explained by prior youth care utilization dummies, by type care received (Van de Ven et al., 2024).

$$(3)\quad ZvwCost_i = \alpha + \sum_{k=1}^{K} \beta_k YouthCareType_{ik} + \varepsilon_i$$

Equation 3 represents the baseline OLS regression model, with dependent variable Zvw costs, and independent variables the different types of youth care received in the prior year. The intercept alpha indicates the mean zvw costs for individuals without prior youth care use. The error term indicates the unexplained variation in costs at the individual level.

3.2.3 Predictive Power for Residual Zvw Costs

The third research question examines to what extent prior use of youth care is predictive of residual health care spending after applying the 2026 RE model. To answer this question, the analysis will compute the mean residual cost respectively for somatic and GGZ care. According to theory, residual costs are a direct indicator of whether certain subgroups are under- or overcompensated, as argued by Van Kleef & Van Vliet (2025). Residual costs are calculated as actual Zvw costs minus predicted Zvw costs from the 2026 RE model, as shown in equation 4.

$$(4)\ ResidualCost_i = ActualZvwCost_i - PredictedZvwCost_i$$

The interpretation of the residual cost indicates that a positive sign indicates undercompensation, meanwhile a negative sign indicates overcompensation (Van Kleef & Van Vliet, 2025). Zero indicates that the subgroup falls in neither of the cases. Next, the analysis also includes two OLS regression analysis, to measure the predictive power of youth care use in the prior year on residual costs. Hence, the analysis will include one base model focusing on costs from the year prior, and one extended model accounting for the costs in year t-2 and t-3.

$$(5)\ ResidualCost_i = \alpha + \sum_{K=1}^{K} \beta_k\, YouthcareType_{ik} + \varepsilon_i$$

Equation 5 represents the baseline OLS regression model, with dependent variable residual Zvw costs, and as independent variables, the different types of youth care received in the prior year. The intercept alpha indicates the mean zvw costs for individuals without prior youth care use. The error term indicates the unexplained variation in costs at the individual level.

3.2.4 Qualitative Assessment of Risk Adjuster Criteria

This section aims to research to what extent prior use of youth care fulfils the criteria for risk adjusters introduced by the WOR framework and Van de Ven & Ellis (2000). In the results section, the most important criteria will be used to assess the previous empirical results. This means that appropriateness of incentives, fairness and feasibility will each be given a score, either positive, if it is proven that 'prior youth care use' adheres to all of the conditions for the criteria to be met. Similarily, it will receive a negative sign, if proven other. Lastly, the criterion can also have a mixed '+/-' sign, if proven that the criteria can only be partially met.

3.2.5 Validity and reliability

This section refers to the validity and reliability of the study design and analytical choices made in this study. The internal validity aims to look at whether the analysis measures what it claims to measure. In this study, coupling the RIN together with youth care data and Zvw costs allows to combine two relevant datasets to analyze the underlying subquestions. In addition, the formulas used in this research are a direct replication of what the subquestions aim to research. For instance, the residual costs are calculated by subtracting actual minus predicted costs from the 2026 RE model, reflecting the third research question on whether youth care has predictive power on residual Zvw costs.

Next, external validity refers to the generalizability of the findings. The dataset covers national cost data on 4.9 million inhabitants under 25, so they are generalizable within the Dutch context. However, the findings of this research are restricted to one reference year, being 2023. Hence, generalizability to other countries or time periods cannot be conducted.

Third, construct validity refers to whether the variable represents the measure what it claims to measure. In the research question, the focus of the analysis lies on 'prior youth care use'. The methodology and analysis steps include year dummies, which are variables on prior youth care use in 2021 and 2020. In addition, the first descriptive analysis has been done with prior year youth care usage, 2022. This confirms that the variable validates for construct validity.

Reliability refers to whether the analysis would repeat consistent results if repeated. Within the context of youth care, the data might be subject to municipal registration inconsistencies, indicating that replicating the data would not necessarily result in the consistency of the results.

Ethics in this study refers to the preservation of privacy of the individuals in the dataset. Since the thesis runs a national dataset base, the RIN identifier ensures that the data of the individual is anonymous. However, this study also acknowledges that the datasets contain sensitive information on health and socioeconomic status. Lastly, the data had been granted only through exclusive access to the CBS microdata agreement, indicating that the dataset is protected.

4. Results

This section discusses the findings of the empirical results, conducted on Rstudio. First, section 4.1 discusses the findings on characteristics of youth care users. Section 4.2 discusses the findings for the predictive power for total zvw costs. Section 4.3 discusses the predictive power of residual Zvw costs, followed by section 4.4, which is the qualitative assessment of the criterion on the fit of 'prior use of youth care' as risk adjuster.

4.1 Characteristics of youth care users

Table 3 presents the descriptive statistics for the individuals under 25 years of age in the population (n= 4,877,025). Of the 4,877,025 individuals in the dataset, 447,676 had prior youth care in 2022. First, the results for somatic risk adjusters indicate that Multiple-Year High cost (MYHC) are large for both groups, with non-youth care showing 31.86% of having at least one Multiple-Year High cost (MYHC). For youth care users, 48.62% had at least one MYHC, indicating that almost half of youth care users in 2022 have a history of high costs. Next, the results show that youth care users are higher consumers of somatic care, indicating that 6.55% of youth care users had at least one Pharmacy-based cost (PCG), and 5.07% had at least one Diagnostic-based cost (DCG).

Surprisingly, youth care users are lower users of morbidity-based care, with 2.60% having at least one PCG in the prior year, contrasting to non-youth care users with 2.65%. The population of non-youth care users is over 4 million, so there might be some unexplained variation in different types of care taken outside youth care within that population. Next, the result for socioeconomic status shows that 30.72% of youth care users have very low socioeconomic status. This adds an additional layer to the claims that youth care users are systematically worse off, on top off them being higher users of somatic care.

Table 3. *Characteristics of youth care users and non-users under age 25*

Variable	No youth care 2022	Youth care 2022
N observations	4,429,349	447, 676
Mean age (years)	12.6	12.2
Risk Adjusters as a % for prevalence rate		
Somatic risk adjusters		
At least one PCG	3.41	6.55
At least one DCG	2.31	5.07
At least one MYHC	31.86	48.62
Mental risk adjusters		
At least one DCG in prior year	2.65	2.60
At least one PCG in prior year	1.00	1.86
At least one MYHC in prior year	3.80	2.12
Other equalisation criteria		
Physiotherapy Diagnosis Group	3.64	9.56
MYNC	0.27	0.78
Socioeconomics status		
Very low	22.52	30.72

Low	18.71	20.31
Middle	28.13	25.12
High	30.64	23.86
Persons per Address		
Reference	66.70	87.35
One person	1.70	0.39
Other household type	31.18	11.72
Wlz institutional care, below threshold	0.13	0.14
Wlz non- institutional care, below threshold	0.26	0.28
Wlz institutional care, above threshold	0.0042	0.05
Wlz non- institutional care, above threshold	0.002	0.07

Note. N = 4,877,025 except for the variable age (n = 4,877,007). The population covered individuals under 25, all risk adjusters are denoted in proportions (%), because their values are binary in the dataset. Reference year 2022 - 2023. Variable names: PCG = Pharmacy-based cost group, DCG = Diagnostic-based cost group ; MYHC = Multiple-Year High Costs group ; mPCG = Multiple-year high costs, pharmacy-based cost group ; pPCG = Pharmacy-based cost group in prior year ; pDCG = Diagnosis-based cost group in prior year ; pMYHC = Multiple-year high cost group ; PDG = physiotherapy diagnosis group ; MYNC = Multiple-year nursing and personal care costs group. Persons per address (PPA) refers to the amount of people living in one household. In the dataset, the PPA classes are subdivided into multiple categories, PPA_1 – PPA_7, but in this table they have been written fully out to ease the reading flow. 'Reference group' refers to the baseline household type, so typically a multi-person household without any Wlz recipient (Van Kleef & Van Vliet, 2025). One person refers to a household consisting of only one person. Other household type refers to households not fitting in each of the PPA categories, for instance multi-person households. Wlz institutional care refers to households receiving long-term care (Wlz) at a care facility, whereas non-institutional care refers to someone receiving care at home. Below or above the threshold refers to the concentration of Wlz recipients at one address. Thus, below refers to a lesser concentration of individuals receiving long-term care, and above threshold refers to a larger concentration of long-term care.

Table 4 provides the output results for the characteristics of youth care users, based on the type of care they received. Findings show that youth care users have a larger burden of morbidity care, with the highest prevalence rate of 68.33% for users of type 3 having at least one Multiple-Year High Cost (MYHC). In addition, 14.57% of users in the same category have at least one Diagnostic-based cost, and 9.87% have at least one Pharmacy-based cost.

Next, results for physiotherapy diagnosis show higher prevalence rates for youth care users compared to non-users (3.64%). Rather, users of type 3 have a prevalence rate of respectively each 20.64%, and users of type 4 have a prevalence rate of 11.52%. This indicates that people with youth care require more diagnosis, as well as the need for treatments. Next, results for socioeconomic status show a prevalence rate of 43.83% for low socioeconomic status among users of type 1. Similarily, users of type 4 have a prevalence rate of 40.67%. These results are consistent with the previous findings in Table 3, indicating that people using youth care are characterised by systematic socioeconomic restraints, as well as have higher needs for treatments.

Table 4. *Characteristics of youth care users by type of care received (type 1 – type 4)*

Variable	No youth care	Type 1: Youth care without residential stay, provided by a district or neighbourhood team	Type 2: Youth care without residential stay, outpatient care at the provider's location	Type 3: Youth care without residential stay, day care at the provider's location	Type 4: Youth care without residential stay, care provided within the young person's social network
N observations	4,429,349	64,095	325,410	34,475	99,566
Mean_age_years	12.60	11.73	12.44	10.09	11.55
Risk Adjusters as a % for prevalence rate					
Somatic risk adjusters					
At least one PCG	3.41	6.67	6.81	9.87	7.26
At least one DCG	2.31	6.06	4.48	14.57	6.27
At least one MYHC	31.86	51.02	48.28	68.33	53.16
Mental risk adjusters					
At least one DCG in prior year	2.65	2.14	2.68	0.97	1.62
At least one PCG in prior year	1.00	1.25	2.04	1.3	1.54
At least one MYHC in prior year	3.80	2.12	1.9	0.76	1.32
Other equalisation criteria					
Physiotherapy Diagnosis Group	3.64	9.73	9.34	20.64	11.52
MYNC	0.27	1.07	0.65	2.33	1.07
Socioeconomic status					
Very low	22.52	43.83	26.14	40.23	40.67
Low	18.71	20.06	20.29	22.4	21.51
Middle	28.13	19.72	27	22.08	21.41
High	30.64	16.38	26.57	15.3	16.42
Persons per Address					
Reference group	66.70	86.79	88.61	93.49	89.82
One person	1.70	0.47	0.22	0.12	0.32
Other household type	31.18	12.16	10.85	5.31	9.33
Wlz institutional care, below threshold	0.13	0.13	0.07	0.15	0.1

Wlz non- institutional care, below threshold	0.26	0.3	0.17	0.59	0.29
Wlz institutional care, above threshold	0.42	0.06	0.03	0.1	0.04
Wlz non- institutional care, above threshold	0.02	0.09	0.05	0.23	0.11

Note. N = 4,877,025 except for the variable age (n = 4,877,007). The population covered individuals under 25, all risk adjusters are denoted in proportions (%), because their values are binary in the dataset. Reference year 2022 - 2023. Variable names: PCG = Pharmacy-based cost group, DCG = Diagnostic-based cost group ; MYHC = Multiple-Year High Costs group ; mPCG = Multiple-year high costs, pharmacy-based cost group ; pPCG = Pharmacy-based cost group in prior year ; pDCG = Diagnosis-based cost group in prior year ; pMYHC = Multiple-year high cost group ; PDG = physiotherapy diagnosis group ; MYNC = Multiple-year nursing and personal care costs group. Persons per address (PPA) refers to the amount of people living in one household. In the dataset, the PPA classes are subdivided into multiple categories, PPA_1 – PPA_7, but in this table they have been written fully out to ease the reading flow. 'Reference group' refers to the baseline household type, so typically a multi-person household without any Wlz recipient (Van Kleef & Van Vliet, 2025). One person refers to a household consisting of only one person. Other household type refers to households not fitting in each of the PPA categories, for instance multi-person households. Wlz institutional care refers to households receiving long-term care (Wlz) at a care facility, whereas non-institutional care refers to someone receiving care at home. Below or above the threshold refers to the concentration of Wlz recipients at one address. Thus, below refers to a lesser concentration of individuals receiving long-term care, and above threshold refers to a larger concentration of long-term care.

Table 5 provides the output results for the characteristics of youth care users, providing the second half of subgroups. More specifically, from type 11 to type 14. Similar to other subgroups of youth care in Table 4, there is a higher need for somatic care, with the highest prevalence rate of 64.71% for having at least one Multiple-Year high Cost (MYHC) for users of type 14. Users of type 12 have a prevalence rate of 6% with at least one Diagnostic-cost group (DCG), and users of type 14 have a prevalence rate of 13.32% with at least one Pharmacy-Cost group (PCG). Again, these results are higher than non-youth care users, for instance 31.86% for having at least one MYHC, indicating that youth care are repeatedly characterized by higher needs of treatments and higher historic costs. Similarly, prevalence rates for a very low socioeconomic status are the highest across youth care users, with highest showing 59.41% and 56.4% for users of type 13 respectively users of type 14. These results are consistent with the previous findings in Table 3 and other subgroups from Table 4, indicating that people using youth care are also characterised by having systematic socioeconomic restraints.

Table 5. *Characteristics of youth care users by type of care received (type 11 – type 14)*

Variable	No youth care	Type 11: Youth with residential stay, foster care	Type 12: Youth with residential stay, family-based care	Type 13: Youth with residential stay, secure residential placement	Type 14: Youth with residential stay, other than the categories listed above
N observations	4429349	20156	6523	1853	18804
Mean_age_years	12.60	12.46	12.96	16.19	14.5
Risk Adjusters as a % for prevalence rate					
Somatic risk adjusters					

At least one PCG	3.41	5.48	9.45	23.7	13.32
At least one DCG	2.31	4.09	6	4.48	5.78
At least one MYHC	31.86	42.48	50.53	64.71	56.85
Morbidity-based risk adjusters					
At least one DCG in prior year	2.65	2.99	3.58	7.36	6.06
At least one PCG in prior year	1.00	1.47	2.92	7.84	4.49
At least one MYHC in prior year	3.80	3.49	3.45	4.13	4.75
Other equalisation criteria					
Physiotherapy Diagnosis Group	3.64	7.12	8.26	3.46	7.01
Multiple-year High cost	0.27	0.54	1.1	1.85	1
Socioeconomic status					
Very low	22.52	39.72	47.02	59.41	56.4
Low	18.71	21.07	13.71	12.48	15.84
Middle	28.13	21.03	17.32	13.28	14.35
High	30.64	18.17	21.95	14.82	13.41
Persons per Address					
Reference	66.70	78.9	80.47	73.62	72.85
One person	1.70	0.7	1.4	1.17	3.38
Other household type	31.18	19.12	15.83	22.53	20.81
Wlz institutional care, below threshold	0.13	0.32	0.73	1.02	1.3
Wlz non- institutional care, below threshold	0.26	0.73	0.88	0.22	0.73
Wlz institutional care, above threshold	0.42	0.07	0.24	0.66	0.61
Wlz non- institutional care, above threshold	0.02	0.16	0.45	0.77	0.31

Note. N = 4,877,025 except for the variable age (n = 4,877,007). The population covered under 25, all risk adjusters are denoted in proportions (%), because their values are binary in the dataset. Reference year 2022 - 2023. Variable names: PCG = Pharmacy-based cost group, DCG = Diagnostic-based cost group ; MYHC = Multiple-Year High Costs group ; mPCG = Multiple-year high costs, pharmacy-based cost group ; pPCG = Pharmacy-based cost group in prior year ; pDCG = Diagnosis-based cost group in prior year ; pMYHC = Multiple-year high cost group ; PDG = physiotherapy diagnosis group ; MYNC = Multiple-year nursing and personal care costs group. Persons per address (PPA) refers to the amount of people living in one household. In the dataset, the PPA classes are subdivided into multiple categories, PPA_1 – PPA_7, but in this table they have been written

fully out to ease the reading flow. 'Reference group' refers to the baseline household type, so typically a multi-person household without any Wlz recipient (Van Kleef & Van Vliet, 2025). One person refers to a household consisting of only one person. Other household type refers to households not fitting in each of the PPA categories, for instance multi-person households. Wlz institutional care refers to households receiving long-term care (Wlz) at a care facility, whereas non-institutional care refers to someone receiving care at home. Below or above the threshold refers to the concentration of Wlz recipients at one address. Thus, below refers to a lesser concentration of individuals receiving long-term care, and above threshold refers to a larger concentration of long-term care.

4.2 Predictive power for total zvw costs

Table 6 presents the mean Zvw costs by type of care received. On an aggregate level, mean total Zvw costs are higher than non-youth care, namely €2383 and €1494. More specifically, mean somatic healthcare costs are prominently higher than mean mental healthcare costs (GGZ), namely €2075 and €340 each. When separating between different care types, users of type 13 have significant total Zvw costs, nearly €8380, as well as high costs for somatic care, € 3818 and €4634 for mental healthcare cost. A similar pattern occurs for users of type 14, with a total Zvw cost of € 4046, and a mean somatic cost of €2460, followed by €1655 for mental healthcare costs.

Table 6. *Mean Zvw costs and mean costs by youth care status and type of care received*

Group	Mean total Zvw costs (€)	Mean Somatic healthcare costs (€)	Mean mental healthcare costs (€)
Reference & Aggregate			
No youth care	1494	1378	172
Youth care (all)	2383	2075	340
By care type (t-1)			
Type 1: Youth care without residential stay, provided by a district or neighbourhood team	2783	2521	292
Type 2: Youth care without residential stay, outpatient care at the provider's location	2197	1869	358
Type 3: Youth care without residential stay, day care at the provider's location	4553	4350	219
Type 4: Youth care without residential stay, care provided within the young person's social network	2677	2416	285
Type 11: Youth care with residential stay, foster care	1868	1620	291
Type 12: Youth care with residential stay, family-based care	2960	2259	748

Type 13: Youth care with residential stay, secure residential placement	8380	3818	4634
Type 14: Youth care with residential stay, other than the categories listed above	4046	2460	1655

Note. N observations = 4,877,025, with reference year 2022-2023. Costs have been rounded to full euro's. For purposes of the analysis, all costs in the file have been converted to an annual basis. For instance, for an insured person with actual costs of 2,000 euros and an enrollment period of 6 months (=0.5 years), the costs have been increased to 2,000/0.5 = 4,000 euros. In short: the file therefore contains the costs per insured year. A weight ("gew") has also been included in the file. This variable represents the fraction of the year that an insured person was enrolled with an insurer in 2023.

Other results show that for users of type 3, somatic costs are the highest, nearly €4350, and are simultaneously the lowest for mental healthcare costs, respectively €219. This indicates that youth care users from this subgroup have higher costs on care revolving around physical needs, rather than mental health.

Table 7 provides the OLS regression output of total Zvw costs based off prior youth care use. The intercept reflects the mean expected Zvw costs for the reference group, being non-youth care users. The cost coefficients at the intercept amount €1388 for somatic care and €172 for mental healthcare costs. Every other coefficient is interpreted relative to the reference group, as additional expected costs associated with each subgroup of youth care use. The highest coefficient for somatic care is for users of type 3, at €2531. This indicates that relative to non-users of youth care, this group has €2531 higher somatic costs. Within that same care type, the coefficient for mental healthcare costs is -€116, indicating that this mean group has lower mental healthcare costs than non-youth care users. Users of type 11 have for both somatic and GGZ negative coefficients, respectively -€211 and -€44. This indicates that this subgroup costs less than the reference group. Users of type 13, which has both positive and elevated costs for both models, each €1495 for somatic and €3733 for mental healthcare, have higher costs relative to non-users.

Table 7. *Base model (OLS) regression with care type of total Zvw costs on prior youth care type*

Variable	Somatic healthcare costs(€)	Mental healthcare costs(€)
Intercept (no youth care)	1388	172
Type 1: Youth care without residential stay, provided by a district or neighbourhood team	747	25
Type 2: Youth care without residential stay, outpatient care at the provider's location	224	144
Type 3: Youth care without residential stay, day care at the provider's location	2531	-116
Type 4: Youth care without residential stay, care provided within the young person's social network	502	-33
Type 11: Youth care with residential stay, foster care	-211	-44
Type 12: Youth care with residential stay, family-based care	221	222
Type 13: Youth care with residential stay, secure residential placement	1495	3733
Type 14: Youth care with residential stay, other than the categories listed above	351	1222
R^2	0.000835	0.001853

Note. N observations = 4,877,025, with reference year 2022-2023.Dependent variable: total Zvw costs (€), distincting between the sub-models, somatic and mental healthcare costs. The dummies in this table are the different subgroups of youth care services refer to the (prior use of youth care in t-1, or 2022). Costs have been rounded to full euro's. For purposes of the analysis, all costs in the file have been converted to an annual basis. For instance, for an insured person with actual costs of 2,000 euros and an enrollment period of 6 months (=0.5 years), the costs have been increased to 2,000/0.5 = 4,000 euros. In short: the file therefore contains the costs per insured year. A weight ("gew") has also been included in the file. This variable represents the fraction of the year that an insured person was enrolled with an insurer in 2023.

The R-square for both sub-models is 0.000835 for somatic care, and 0.001853 for mental healthcare. For somatic care, this means that prior youth care explains 0.08% of the variance in Zvw costs. Mental healthcare, which has stronger predictive power due to the higher R-square, indicates that prior youth care explains 0.19% of the variance in mental healthcare costs. These results show that prior youth care use has explanatory power, but in absolute terms the value remains very small. This means that there are other factors potentially contributing to the variation in costs for Zvw.

Table 8 represents the extended model version of the OLS regression results of the total Zvw costs on prior youth care type. In this model, 2 additional year dummies are added, being previous youth

care use in the year 2021 and 2020, to test whether looking at historical youth care trajectory has an impact on current spending. The intercept reflects the mean expected Zvw costs for the reference group, being non-youth care users. The cost coefficients at the intercept amount €1374 for somatic care and €150 for mental healthcare costs. Every other coefficient is interpreted relative to the reference group, as additional expected costs associated with each subgroup of youth care use. The results of this model have similar coefficient signs as the base model, with the coefficient of somatic costs being €2444, indicating that relative to non-users of youth care, users of care in their own network have higher somatic costs. Similar to the base model, a negative cost coefficient for mental healthcare, -€243, indicates that relative to non-users of youth care, this group has lower somatic care costs.

The findings of the year coefficients indicate that youth care use in 2021 are €156 higher compared to the reference group of non-users for somatic care, and respectively €79 for mental care. Next, youth care use in 2020 shows an increase of €157 for somatic and €401 for GGZ. For somatic care, the costs didn't change that much, but for mental healthcare costs, the increase over time is quite notable. This suggests that tracking back 3 years prior of historical youth care trajectory adds meaningful information for mental healthcare costs.

The R-square for both sub-models is 0.000876 for somatic care, and 0.003328 for mental healthcare. For somatic care, this means that prior youth care explains 0.08% of the variance in Zvw costs. Mental healthcare, which has stronger predictive power due to the higher R-square, indicates that prior youth care explains 0.33% of the variance in mental healthcare costs. These results show that prior youth care use has explanatory power, but in absolute terms the value remains very small. Compared to the base model, the R-square did not increase significantly for the somatic model, but for the mental health costs, increased by +0.14%. This shows that although in absolute terms the R-square remains small, adding year dummies, that tracked use of youth care for a longer period, has increased its explanatory power for that sub-model.

Table 8. *Extended OLS regression model of total Zvw costs on prior youth care type*

Variable	Somatic healthcare costs (€)	Mental healthcare costs(€)
Intercept (no youth care)	1374	150
Type 1: Youth care without residential stay, provided by a district or neighbourhood team	646	-113
Type 2: Youth care without residential stay, outpatient care at the provider's location	97	-12
Type 3: Youth care without residential stay, day care at the provider's location	2444	-243
Type 4: Youth care without residential stay, care provided within the young person's social network	398	-173
Type 11: Youth care with residential stay, foster care	-389	-320
Type 12: Youth care with residential stay, family-based care	108	43
Type 13: Youth care with residential stay, secure residential placement	1472	3694
Type 14: Youth care with residential stay, other than the categories listed above	240	1044
Year Dummies		
Previous use of youth care (2021)	155	79
Previous use of youth care (2020)	157	401
R^2	0.000876	0.003328

Note: N observations = 4,877,025, with reference year 2022-2023. Dependent variable: total Zvw costs (€), distincting between the sub-models, somatic and mental healthcare costs. The dummies in this table are the different subgroups of youth care services refer to the (prior use of youth care in t-2 and t-3, or 2021 and 2020). The year dummies are set as binary variables (Yes/no). Costs have been rounded to full euro's. For purposes of the analysis, all costs in the file have been converted to an annual basis. For instance, for an insured person with actual costs of 2,000 euros and an enrollment period of 6 months (=0.5 years), the costs have been increased to 2,000/0.5 = 4,000 euros. In short: the file therefore contains the costs per insured year. A weight ("gew") has also been included in the file. This variable represents the fraction of the year that an insured person was enrolled with an insurer in 2023.

4.3 Predictive power of residual Zvw costs

Table 9 represents the output results for the mean residual Zvw costs by youth care type received. On average, the results of the reference group, being non-youth care users, indicates that the current risk equalisation model undercompensates somatic care by €68, and overcompensates mental healthcare by -€12. The results for each of the subgroups have positive residuals, for each sub-model too. This indicates that each of the sub-groups are systemically undercompensated by the current 2026 RE model.

Table 9. *Mean residual Zvw costs by youth care group in the previous year (2022)*

Youth care type	Mean residual costs for somatic care(€)	Mean residual costs for mental healthcare (€)
No youth care (reference group)	68	-12
Type 1: Youth care without residential stay, provided by a district or neighbourhood team	198	118
Type 2: Youth care without residential stay, outpatient care at the provider's location	124	198
Type 3: Youth care without residential stay, day care at the provider's location	372	120
Type 4: Youth care without residential stay, care provided within the young person's social network	195	152
Type 11: Youth care with residential stay, foster care	11	90
Type 12: Youth care with residential stay, family-based care	107	461
Type 13: Youth care with residential stay, secure residential placement	1302	3404
Type 14: Youth care with residential stay, other than the categories listed above	232	1010

Note. The formula used to compute the results is actual costs minus predicted costs, with a positive residual indicating an undercompensation, and a negative residual indicating overcompensation. N observations = 4,877,025, with reference year 2022-2023. The dummies in this table are the different subgroups of youth care services refer to the (prior use of youth care in t-1, or 2022). Costs have been rounded to full euro's. For purposes of the analysis, all costs in the file have been converted to an annual basis. For instance, for an insured person with actual costs of 2,000 euros and an enrollment period of 6 months (=0.5 years), the costs have been increased to 2,000/0.5 = 4,000 euros. In short: the file therefore contains the costs per insured year. A weight ("gew") has also been included in the file. This variable represents the fraction of the year that an insured person was enrolled with an insurer in 2023.

Users of type 11 show the lowest undercompensation for both models, with on average €11 for somatic care, and €90 for mental healthcare. This indicates that the RE model performs relatively well for compensating this subgroup. Other results show that the highest mean residual undercompensation is for users of type 13, with mean costs of €1302 for somatic care, and €3404 for mental healthcare. This indicates that the Risk-equalisation model severly undercompensates insurers for this subgroup, leading to predictable losses as well. Next, users of type 14 have a high mean undercompensation for mental healthcare costs, €1010, and a lower undercompensation for somatic care, €232. The difference in costs indicates that this subgroup incurs higher mental healthcare costs, yet the RE model fails to compensate adequately for it, compared to somatic care. Overall, these results show an undercompensation for each of the subgroups, with largest burden on insurers for users of type 13 and 14.

The results presented in Table 10 provide the results of the OLS regression of residual Zvw costs on prior youth care type subgroups, estimated separately for the two sub-models of the Risk equalization model (RE). The intercept, which reflects the residual costs for non-youth care users, is

€71 for somatic care, indicating that there is a slight undercompensation for somatic care. For mental healthcare, the coefficient is -€12, indicating a slight overcompensation for mental healthcare cost. Next, the coefficient results for users of type 3 is €247 for somatic care, indicating an additional undercompensation relative to the reference group. For mental healthcare, the coefficient result is -€34, indicating even higher overcompensation compared to non-youth care users. Next, users of type 11 negative coefficients for both sub-models with -€111, showing a higher overcompensation relative to non-users, which even shows an undercompensation (€71). For mental healthcare, the coefficient is -€45.64, indicating a relatively higher overcompensation relative to non-users. In this regression, this subgroup is the only type to show overcompensation for both submodels. Users of type 13, which shows the largest coefficients in both submodels, has an undercompensation of €1134 for somatic care, and an undercompensation of €2880 for mental healthcare, for both relative to non-users. This indicates that the RE model fails to compensate adequately for this subgroup both in terms of somatic and mental healthcare.

Table 10. *OLS regression of residual Zvw costs on prior youth care type in the previous year (2022)*

Variable	Somatic healthcare costs - coefficient (€)	Mental healthcare costs - coefficient (€)
Intercept (no youth care)	71	-12
Type 1: Youth care without residential stay, provided by a district or neighbourhood team	84	28
Type 2: Youth care without residential stay, outpatient care at the provider's location	24	174
Type 3: Youth care without residential stay, day care at the provider's location	247	-34
Type 4: Youth care without residential stay, care provided within the young person's social network	67	28
Type 11: Youth care with residential stay, foster care	-111	-45
Type 12: Youth care with residential stay, family-based care	-44	186
Type 13: Youth care with residential stay, secure residential placement	1134	2880
Type 14: Youth care with residential stay, other than the categories listed above	44	767
R^2	0.000019	0.001577

Note. N observations = 4,877,025, with reference year 2022-2023. Dependent variable: total Zvw costs (€), distincting between the sub-models, somatic and mental healthcare costs. The dummies in this table are the different subgroups of youth care services refer to the (prior use of youth care in t-1, or 2022). Costs have been rounded to full euro's. For purposes of the analysis, all costs in the file have been converted to an annual basis. For instance, for an insured person with actual costs of 2,000 euros and an enrollment period of 6 months (=0.5 years), the costs have been increased to 2,000/0.5 = 4,000 euros. In short: the file therefore contains the costs per insured year. A weight ("gew") has also been included in the file. This variable represents the fraction of the year that an insured person was enrolled with an insurer in 2023.

Next, users of type 14 have a coefficient of €44 for somatic care, and has largely higher coefficient for mental healthcare, namely €767.

The R-square of this model shows very small results for both sub-models, namely 0.000019 for somatic and 0.001577. For somatic care, this means that prior youth care explains 0.0019% of the variance in Zvw costs, which is increasingly small compared to previous R-square values in previous models (0.0019% in the base model Table 6). Mental healthcare, which has stronger predictive power due to the higher R-square, indicates that prior youth care explains 0.16% of the variance in mental healthcare costs. The low R-square values indicate that prior youth care use has limited explanatory power at the individual level, even after accounting for residual compensation.

Table 11 provides the extended model regression for residual Zvw costs on prior youth care type for 2021 and 2020. The intercept of this model shows for somatic care that non-youth care users are undercompensated by €74. For mental healthcare, non-youth care users show an overcompensation of- €19. For ussers of type 4, the somatic coefficient is €91, showing an under compensation relative to non-users. For mental healthcare, the coefficient is – €15, showing relative overcompensation for mental healthcare. Next, the result coefficients for users of type 11 are – €68 for somatic care, showing an overcompensation relative to non-users. For mental health care, the coefficient is - €128, indicating a larger overcompensation relative to non-users. Larger differences in costs within subgroups could be due to higher mental healthcare treatments compared to somatic care.

Next, the largest residual coefficients are for users of type 13, with somatic care €1140, showing a large undercompensation relative to non-users. For mental healthcare, the coefficient is €2869, indicating a large undercompensation relative to non-users. With the current RE model not yet accounting for youth care, the residual over- and undercompensations show that there is no adequate integration of youth care to adjust for these differences. The coefficient results for 2021 amount up to -€29 for somatic care, and €41 for mental health care. In 2020, the result coefficients are each €45 for somatic care, and €103 for mental health care. Both year dummies indicate historical use of youth care is linked to lower somatic residuals, but higher residuals for mental health care. Repeatedly, the results indicate that historical trajectories of 3 years prior are particularly relevant for mental healthcare undercompensation.

The R-square of this model are slightly higher than the base residual model (0.0019% for somatic and 0.15% for mental healthcare). For somatic care, this value is 0.000022, and for mental health care, 0.001745. For somatic care, this means that prior youth care explains 0.0022% of the variance in Zvw costs. For mental healthcare, which has stronger predictive power due to the higher R-square, indicates that prior youth care explains 0.17% of the variance in mental healthcare costs. The low R-square values indicate that prior youth care use has limited explanatory power at the individual level, even after accounting for residual compensation.

Table 11. *OLS regression of residual Zvw costs on prior youth care type (2021, 2020)*

Term	Somatic healthcare costs - coefficient (€)	Mental healthcare costs - coefficient (€)
Intercept (No youth care)	74	- 19
Type 1: Youth care without residential stay, provided by a district or neighbourhood team	107	- 14
Type 2: Youth care without residential stay, outpatient care at the provider's location	52	123
Type 3: Youth care without residential stay, day care at the provider's location	267	- 73
Type 4: Youth care without residential stay, care provided within the young person's social network	91	- 15
Type 11: Youth care with residential stay, foster care	- 68	- 128
Type 12: Youth care with residential stay, family-based care	- 17	133
Type 13: Youth care with residential stay, secure residential placement	1140	2869
Type 14: Youth care with residential stay, other than the categories listed above	71	714
Year dummies		
Prior use of youth care in 2021	- 29	41
Prior use of youth care in 2020	- 45	103
R-squared	0.000022	0.001745

Notes. The formula used to compute the results is actual costs minus predicted costs, with a positive residual indicating an undercompensation, and a negative residual indicating overcompensation. N observations = 4,877,025, with reference year 2022-2023. The dummies in this table are the different subgroups of youth care services refer to the (prior use of youth care in t-2 and t-3, or 2021 and 2020). The year dummies are set as binary variables (Yes/no). Costs have been rounded to full euro's. For purposes of the analysis, all costs in the file have been converted to an annual basis. For instance, for an insured person with actual costs of 2,000 euros and an enrollment period of 6 months (=0.5 years), the costs have been increased to 2,000/0.5 = 4,000 euros. In short: the file therefore contains the costs per insured year. A weight ("gew") has also been included in the file. This variable represents the fraction of the year that an insured person was enrolled with an insurer in 2023.

4.4 Qualitative assessment

The fourth research question aims to research whether prior use of youth care fulfils the criteria as a risk adjuster. Table 11 presents the most relevant criteria, based off the existing criterion framework from Van de Ven & Ellis (2000) and the WOR are used to interpret the empirical results (Van Kleef et al., 2023). Each criterion takes either a positive or negative sign, depending on whether prior youth

care use satisfies the condition. The appropriateness of incentives refers to whether the risk adjuster creates the right financial incentive for insurers (Van de Ven & Ellis, 2000). The empirical results have shown that youth care users are undercompensated across all subgroups. Integrating youth care as a risk adjuster would absorb the undercompensation faced by insurers, hence reducing incentives for cream-skimming (Van Kleef & Van Vliet, 2025). In addition, since youth care is administered through municipalities and not the Zvw, existing data remains independent from the influence of insurers to manipulate the data and stimulate market inefficient behaviour.

Table 12. *Qualitative assessment on the fit of youth care as risk adjuster*

Criterion	**Assessment**
Appropriateness of incentives	+
Fairness	+
feasibility	+/-

Next, the fairness criterion refers to the discussion on whether cost differences associated with a variable are due to N-type (need-based) or S-type (efficiency-driven) factors. Central to the criterion of fairness lies whether society should subsidise the cost differences for youth care. Clearly, youth care reflects N-type factors. The empirical results on the characteristics of youth care users show systematically higher rates of morbidity based treatments, as well as profiles issued from lower socioeconomic status. These characteristics are systematically disadvantaging younger people, hence are associated with higher healthcare costs that individuals cannot simply choose to avoid (Netherlands Youth Institute, 2019). Third, feasibility refers to whether the variable can be implemented in the RE-model (Van de Ven & Ellis, 2000). Since the youth care act has only been put in place since 2015, municipal registration practices are still maturing and not yet fully standardised across all Dutch municipalities (CBS, 2025). This introduces data quality concerns that need to be resolved before the variable can be used in the risk equalisation model.

Taken together, these findings suggest that prior youth care use satisfies certain, but not all criteria for integrating it as risk adjuster. Since youth care shows clear confirmation on the appropriateness of incentives and fairness, it indicates that youth care fits in terms of core values and framework. However, the mixed assessment for feasibility indicates that the practical aspect of integrating youth care within the RE framework, requires much more attention to detail, but is crucial for a full transmission of the variable.

5. Discussion and Conclusion

5.1 Summary of main key findings

The goal of this study is to examine the potential of prior use of youth care as risk adjustor within the risk-equalisation model. More specifically, this study aims to give an answer to the following research question : 'To what extent can prior use of youth care be used as a risk adjuster to improve the Dutch risk-equalisation model under the Health Insurance Act (Zvw)?'. The results of this study provide an exploratory analysis for evaluating whether prior youth care use constitutes a relevant, appropriate and feasible risk adjuster within the Dutch Zvw risk-equalisation system.

The first sub question aims to research what the characteristics of youth care users are. The results from the descriptive analysis show that youth care users in the prior year are on average a systematically higher-risk group relative to non-users. This is because across the population, youth care users are characterised with a higher prevalence of multi-year high costs, as well as have higher hospital diagnosis or use of chronic medication than non-care users. Other findings show that youth care users have a higher prevalence of being in a very low socioeconomic status, adding a systematic layer to the profiles of youth care users. Overall, these characteristics indicate that prior youth care use is associated with higher burden of healthcare needs, as well as a socioeconomic disadvantage that might persist in later stages of life.

The second sub question aims to research whether prior youth care use has predictive power on total Zvw costs. The results of step 2, which looked at the mean Zvw costs, indicate that usage of youth care in the prior year (2022) shows a strong linkage with Zvw costs. On average, youth care users cost more than non-users, with variety of costs across different sub-groups. Especially for users of secure residential placement (type 13), who showed elevated costs for both sub-models, with mean costs of €3818 for somatic and €4634 for mental healthcare. In addition, the regression results for the base model confirm that youth care users have higher costs relative to non-users, accounting for different youth care subgroups. Although the R-square of the base model remains low in absolute terms (0.08% for somatic and 0.19% for mental healthcare), results in the extended model have shown that adding dummies about historical youth care trajectory in 2021 and 2020 has improved the predictive power of youth care. Especially for the mental healthcare model, historical youth care trajectory increased its R-square from 0.19% to 0.33%, suggesting a stronger association with future Zvw costs than the somatic care model.

The third sub question aims to research whether prior youth care use has predictive power on residual Zvw costs. The results of the residual costs, which were calculated by subtracting the actual cost minus the predicted costs from the 2026 RE model, with a positive residual indicating an under

compensation, and a negative residual indicating overcompensation. The residual results showed that every subgroup under youth care is undercompensated for by the insurers. For instance, secure residential placement (type 13)showed the largest gap, with under compensation reaching €1,300 for somatic care and €3,400 for mental healthcare. Although the predictive power is low in terms of R-square, the residual results indicating under compensation for all subgroups are valuable, indicating a strong linkage with Zvw costs.

In the extended model, the results have shown that adding dummies on historical youth care trajectory in 2021 and 2020 have improved the predictive power of youth care. Repeatedly, the R-square for mental health care has increased slightly from 0.16% to 0.17%. On the other hand, for somatic care, the year coefficients are negative, indicating that historical youth care use is connected to lower residual costs compared to the reference group. Hence, this indicates that the current Risk equalisation model absorbs for somatic costs, through risk adjusters like PCGs or DCGs. However, the higher residual costs for mental healthcare indicate that there is no equivalent framework absorbing the costs linked to mental healthcare (within the context of youth care). This is an important finding, as this indicates that the predictive power of youth care on residual zvw costs puts into spotlight that there is a missing point for mental healthcare. Within this context, adding prior youth care as a risk adjuster would be of substantial benefit to absorb this asymmetry.

The fourth sub question aims to qualitatively assess the fit of 'prior use of youth care' as risk adjuster. Hence, the qualitative assessment in step 4 evaluates youth care against the criteria established by Van de Ven and Ellis (2000) and the Dutch WOR framework. The appropriateness of incentives criterion is met with a strong sense: the residual results from table 9 – 11 indicated under compensation of each of the subgroups, which directly translated to predictable losses for insurers. Integrating prior use of youth care would absorb these residuals, eliminates the incentives for risk selection and predictable losses on care types (Van de Ven & Ellis, 2000).

The fairness criterion is also met, since the descriptive analysis on characteristics of youth care types have shown that youth care users have low socioeconomic status, and exhibit higher morbidity than non-users. Therefore, taking out youth care reflects an act not chosen out of efficiency by the individual, rather because there is a need for it, rendering the associated cost differences N-type in nature (Van de Ven & Ellis, 2000). The feasibility criterion shows a mixed result, the data on youth care is fully available to use (as has been done in this analysis) and has also been linked to the pseudonymised identifier (rin) to facilitate reading and combining the datasets that formulate the interpretation (CBS, 2025). However, it is important to take into account that variation in the reporting

of municipality practices can introduce heterogeneity across municipalities. Hence, this might pose a limitation to the validity and reliability of the existing data.

The findings of the four parts of the analysis provide an empirical basis for evaluating the fit and predictive power of prior use of youth care as risk adjuster in the Dutch Zvw risk equalization model. The summary of main key findings show that youth care users are a higher need group, which has not yet been recognised by the current risk equalisation model. Next, although the R-square of the models remain small, the predictive power remains substantial on the current undercompensation for each subgroup under youth care, highlighting the current financial losses insurers face for each subgroup. The qualitative assessment indicates that youth care shows conditional readiness for implementation, requiring the need for further development.

5.2 Strengths, Limitations and Research recommendations

The findings of the analysis have brought valuable insights on both the significance and meaning of the predictive power of prior use of youth care as risk adjustor. While this study has many strengths, it is still subject to limitations which are important to take into consideration.

First, this study has many strengths, starting with the national coverage of the data. The used dataset covers 4.9 million individuals, which is the full population of insured individuals under 25 in the Netherlands, meaning that the data are directly applicable to the current risk equalisation system. Next, the datasets were linked via a RIN identifier, allowing for a direct association of the real cost data between Zvw costs and prior youth care use. Third, this study used the historical trajectory of youth care users to ensure predictive validity of the potential new risk adjuster. Using three consecutive years to look at historical usage of youth care shows a robust empirical finding of the current output results.

Fourth, within the analysis, the use of predicted costs from the actual 2026 RE model was used, which allows this research to reflect an analysis done with real cost data, as well as address real-life concerns and observations. Fifth, the existing subgroups based off care type were each taken separately within the analysis and not as a binary variable. By doing this, a more detailed outlook is provided when analysing the characteristics of youth care users, as well as interpreting the heterogeneity in results of the OLS regression results.

This research also included certain limitations that are important to highlight. First, this study restricted the analysis and reference year to only one year (2023). The limitation of doing a cross-sectional analysis is that this study design does not look at specific periods of time for observing different cost patterns, such as in a time-series design. Although this research did look at previous years of youth care utilization, in this research they were set as dummies, and not as reference year. In addition, the risk of employing this study design is that having one reference year might only provide temporary stability of the findings Resultantly, future research could replicate the same analysis by

taking a different design angle, such as employing a time series design, with multiple reference years to test the stability of the coefficients. This could improve the robustness of the findings and the conclusion.

Second, in the analysis of assessing the fit of prior use of youth care as a risk adjuster, this research mainly focused on three criterion, being appropriateness of incentives, feasibility and fairness. However, another crucial criterion like stability, which measures the robustness of youth care has not been evaluated in the qualitative assessment of this research. On top off this, another limitation is that the qualitative assessment in this research is an exploratory analysis, rather than a formal assessment. Therefore, future research could conduct a formal WOR assessment that empirically tests the stability of youth care, alongside the other criteria. Conducting this step is crucial before including youth care formally within the risk equalisation model as risk adjuster. Doing so would allow to derive a clear framework on the readiness of youth care as risk adjuster.

Fourth, within the qualitative assessment, the feasibility criterion received a mixed score. This score was given under the argument that registration across municipalities might be inconsistent. Since the quality of the registers cannot be controlled, the data could be subject to measurement errors, creating reliability issues. Therefore, future research could examine whether these inconsistencies affect the stability of youth care cost coefficients, by comparing results across municipalites. Ideally, one could also implement data quality indicators that track the accuracy of the data registered.

Fifth, various limitations have been discussed prior to the argument of using the R-square as primary measure of predictive power. Since the R-square inflates automatically when adding a variable to an existing framework or regression model, hence not reflecting the true predictive power of youth care. Resultantly, the reported R-squares in the empirical analysis might undervalue the actual relevance of adding prior youth care as a risk adjuster. Therefore, future research could explore alternative quantitative instruments that measure compensation gaps.

Lastly, future research could benefit examining comparable health insurance markets like Germany, Belgium or Switzerland, who might provide insights on how their national health insurance compensations intersect with youth care use. Comparing the similarities and structural differences in youth care systems across these markets might allow to analyse and test the transferability of findings for future research.

5.3 Policy implications

This section is directed towards policymakers like the Ministry of Health, Welfare and Sport, or the WOR, who could make use of this exploratory research on youth care for future risk equalisation research.

First, the findings of this research show that youth care could be integrated conditionally as risk adjuster. The term ‘conditional’ in this context is emphasized because not each of the criteria were met, in addition to not having conducted a formal WOR assessment. However, the empirical findings have proven in consecutive steps that youth care subgroups show systematic undercompensation under the current 2026 RE model. The financial loss that insurers enroll on these subgroups might lead to incentives for risk selection, which could carefully be monitored (Van de Ven et al., 2022). For instance, users in subgroups like youth protection or probation could also receive more attention from policymakers, given that that they are more severely undercompensated than other groups. This is especially relevant given that there is no existing mechanism that absorbs these elevated costs upon turning 18. Therefore, policymakers could consider adding prior youth care use as a formal risk adjuster, to reduce incentives for cream-skimming, as well as allowing a level playing field for insurers.

Second, it could be recommended to standardise the registration of youth care registration across municipalities. The previous limitations raised the issue of possible inconsistencies in registration of youth care data across all Dutch municipalities. Maintaining a follow-up on the registrations through implementing data quality indicators on the registrations across municipalities could be vital to improve the completeness of data registration. However, this requires time and budget to be integrated across the systems, and are therefore relevant to take into consideration in a plan of action.

Third, it could be recommended to monitor the growth in youth care utilization adequately. Youth care, which has grown 23% between 2015 and 2022, indicates that future fluctuations should closely be monitored across registration centers. Monitoring the growth of the variable is important, because if youth care were to keep growing, the risk equalisation model should be able to keep pace with these changing patterns.

Lastly, it could be recommended to initiate a formal WOR assessment on the fit of youth care, in which each of the six criteria are assessed. By evaluating criteria that have not been evaluated within the scope of this research, such as stability and feasibility, the following assessment could provide a comprehensive basis for determening the formal fit of youth care in the risk equalisation model. Overall, the policy implications provide a starting point for further policy and institutional asssessment bodies such as the WOR or Ministry of VWS to explore.

5.4 Conclusion

This study served as an exploratory research on the fit of youth care as potential risk adjuster to improve the risk-equalisation model. The results indicate that prior use of youth care could be a meaningful risk adjuster, because the current risk equalisation model fails to cover insurers adequately

for the existing undercompensation across each subgroup. Including youth care can improve the risk equalization model by reducing the incentives for risk selection, as well as stimulating a level playing field for insurers. However, the current conditions do not allow youth care to formally be admitted as risk adjustor because of the current available information needs to be monitored and standardized effectively. Therefore, ensuring that there is complete and uniform registration of youth care will increase the chances of its integration in the risk equalization model.

6. References

Bakker, R. (2023). *1. Jongeren met jeugdhulp*. Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/longread/rapportages/2023/jeugdhulp-2022/1-jongeren-met-jeugdhulp

Barros, P. P. (2003). *Cream-skimming, incentives for efficiency and payment system*. https://www.sciencedirect.com/science/article/abs/pii/S0167629602001194?via=ihub

Centraal Bureau voor de Statistiek. (2025). *Gebruikershandleiding Zorgregisters: Informatie gebruik*. Centraal Bureau Voor De Statistiek. https://www.cbs.nl/nl-nl/onze-diensten/maatwerk-en-microdata/microdata-zelf-onderzoek-doen/microdatabestanden/gebruikershandleiding-zorgregisters-informatie-gebruik

Enthoven, A. C., & Van De Ven, W. P. (2007). Going Dutch — Managed-Competition health insurance in the Netherlands. *New England Journal of Medicine*, *357*(24), 2421–2423. https://doi.org/10.1056/nejmp078199

Jeurissen, P., & Maarse, H. (2021). *The market reform in Dutch health care: plan and strategy*. The Market Reform in Dutch Health Care - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK577818/

Kroneman, M., & Beerman, B. (2024). *Reform Youth Agenda (Hervormingsagenda Jeugd)*. OBS. https://eurohealthobservatory.who.int/monitors/health-systems-monitor/updates/hspm/netherlands-2016/reform-youth-agenda-(hervormingsagenda-jeugd)

Leu, R. E., Rutten, F. F. H., Brouwer, W., Matter, P., & Rütschi, C. (2009, January). *the swiss and dutch health insurance systems: universal coverage and regulated competitive insurance markets*. https://search.issuelab.org/resources/11352/11352.pdf

Méndez Mediavilla, F. A., Landram, F., & Shah, V. (2002). A comparison of the coefficient of predictive power, the coefficient of determination and AIC for linear regression. *Journal of Applied Business and Economics*. http://www.na-businesspress.com/JABE/MendezWeb.pdf

Ministerie van Volksgezondheid, Welzijn en Sport. (2016, January). *Het Nederlandse zorgstelsel*. https://open.overheid.nl/documenten/ronl-5f6ea9d9-c8b5-4f2a-903b-1732d9579578/file

Ministerie van Volksgezondheid, Welzijn en Sport. (2025, November 18). *Welke jeugdhulp is mogelijk vanuit de Jeugdwet*. Ministerie Van Volksgezondheid, Welzijn En Sport. https://www.regelhulp.nl/onderwerpen/wetten-regels/jeugdwet/wat-is

Nederlandse Zorgautoriteit, & Langejan, T. W. (2012). *Marktscan Mondzorg*. https://zoek.officielebekendmakingen.nl/blg-174955.pdf

Rooijen, M. R., De Jong, J. D., & Rijken, M. (2011). Regulated competition in health care: Switching and barriers to switching in the Dutch health insurance system. *BMC Health Services Research*, *11*(1), 95. https://doi.org/10.1186/1472-6963-11-95

Trescher, A., Listl, S., Van Der Galien, O., Gabel, F., & Kalmus, O. (2019). Once bitten, twice shy? Lessons learned from an experiment to liberalize price regulations for dental care. *The European Journal of Health Economics*, *21*(3), 425–436. https://doi.org/10.1007/s10198-019-01145-z

Van De Ven, W., Hamstra, G., Van Kleef, R., Reuser, M., & Stam, P. (2022). The goal of risk equalization in regulated competitive health insurance markets. *The European Journal of Health Economics*, *24*(1), 111–123. https://doi.org/10.1007/s10198-022-01457-7

Van De Ven, W. P. M. M. (2026). Risk equalization and prevention. *The European Journal of Health Economics*. https://doi.org/10.1007/s10198-026-01935-2

Van De Ven, W. P., & Schut, F. T. (2008). Universal mandatory health insurance in the Netherlands: a model for the United States? *Health Affairs*, *27*(3), 771–781. https://doi.org/10.1377/hlthaff.27.3.771

Van Kleef, R. C., Eijkenaar, F., Van Vliet, R. C., & Van De Ven, W. P. (2018). Health plan payment in the Netherlands. In *Elsevier eBooks* (pp. 397–429). https://doi.org/10.1016/b978-0-12-811325-7.00014-2

Van Kleef, R. C., & Van De Ven, W. P. M. M. (2024). *A Critical review of the use of R2 in risk equalization research*. https://pmc.ncbi.nlm.nih.gov/articles/PMC11937047/

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., R.C.J.A. Van Vliet, & ESHPM, Rotterdam: Erasmus Universiteit. (2026). Verevenende Werking & Power. *WPR-analyse Risicoverevening 2026:* https://www.eur.nl/eshpm/media/2026-05-eindrapport-wpr-analyse-rv26-v30apr2026-1

7. Appendix

Formula's:

(1) $ZvwCost_i = \alpha + \sum_{k=1}^{K} \beta_k YouthCareType_{ik} + \varepsilon_i$

Explanation :

- $ZvwCosti = Actual\ Zvw\ Healthcare\ cost\ for\ individual\ I$
- $\alpha = intercept\ (\mu cost\ for\ reference\ group\ when\ there\ is\ no\ youth\ care)$
- $Youthcare\ Type =$
  $dummy\ variables\ as\ indication\ for\ youth\ care\ in\ the\ previous\ years$
  $(ambulatory, residential, foster\ care), with\ base\ no\ youth\ care$
- $\beta = additional\ expected\ Zvw\ costs\ associated\ with\ youth\ care\ type\ k$
- $\varepsilon_i = error\ term$

(2) $ResidualCost_i = \gamma + \sum_{K=1}^{K} \delta_k\ YouthcareType_{ik} + \varepsilon_i$

Explanation :

- $ResidualCosti =$
  $residual\ Zvw\ healthcare\ cost\ for\ individual\ i\ (i.e.\ the\ unexplained\ part$
  $of\ Zvw\ costs\ after\ accounting\ for\ baseline\ factors\ from\ the\ main\ model)$
- $\gamma = intercept\ (\mu cost\ for\ reference\ group\ when\ there\ is\ no\ youth\ care)$
- $Youthcaretype =$
  $dummy\ variables\ indicating\ exposure\ to\ youth\ care\ type\ k\ in\ previous\ years$
- $(residential, ambulatory, foster\ care), with\ no\ youth\ care\ as\ the\ reference\ category)$
- $\varepsilon_i = error\ term$

Tables:

Table 1. *All variables within the dataset*

Variable name	Description
rin	Anonymous personal identifier (CBS)
leeftijd	Age in years
gew	Sampling weight
mfk	Multiple-Year High Costs group (risk adjuster)

Kost_zvw	Actual total Zvw healthcare costs
Ydak_zvw	Predicted total Zvw costs (risk equalisation model)
Res_zvw	Residual = actual – predicted Zvw costs
Kost_som	Actual somatic care costs
Ydak_som	Predicted somatic costs
Kost_er	Actual first-line/GP care costs
Ydak-er	Predicted first-line costs
Kost_ggz	Actual mental health care costs
Ydak_ggz	Predicted mental health costs
pfkg	Pharmacy Cost Group – chronic medication use (0/1)
pdkg	Diagnosis Cost Group – hospital diagnoses (0/1)
pmhk	Multiple-year high costs pharmacy group (0/1)
mvv	Multiple-year nursing and personal care cost group (0/1)
In_jeugdwet	Youth care user prior year (1=yes, 0=no)
Jgdhulp2022	Youth care use in 2022 (binary)
Jgdhulp2021	Youth care use in 2021 (binary)
Jgdhulp2020	Youth care use in 2020 (binary)
Hulp_01	Youth care type 1 (1=yes)
Hulp_02	Youth care type 2 (1=yes)
Hulp_03	Youth care type 3 (1=yes)
Hulp_04	Youth care type 4 (1=yes)
Hulp_11	Youth care type 11 (1=yes)
Hulp_12	Youth care type 12 (1=yes)
Hulp_13	Youth care type 13 (1=yes)
Hulp_14	Youth care type 14 (1=yes)
PPA_1	Persons per addres – reference group (0/1)
PPA_2	Persons per addres – one person (0/1)
PPA_3	Persons per addres – other household type (0/1)
PPA_4	Persons per addres – Wlz institutional care, below threshold (0/1)
PPA_5	Persons per addres – Wlz non-institutional care, below threshold (0/1)
PPA_6	Persons per addres – Wlz institutional care, above threshold (0/1)
PPA_7	Persons per addres – Wlz non-institutional care, above threshold (0/1)

Note. This table contains the main abbrevations used from the dataset.