



# Do we care about high-cost patients? Estimating the savings on health spending by integrated care

Karen Geurts<sup>1,3</sup> · Marc Bruijnzeels<sup>2</sup> · Erik Schokkaert<sup>3</sup>

Received: 3 April 2021 / Accepted: 7 January 2022 / Published online: 25 January 2022  
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

A recent integrated health care initiative in Belgium supports 12 regional pilot projects scattered across the country and representing 21% of the population. As in shared savings programs, part of the estimated savings in health spending are paid out to the projects to reinvest in new actions. Short-term savings are expected in particular from cost reductions among high-cost patients. We estimate the effect of the projects on spending using a difference-in-difference model. The sensitivity of the results to the right-skewness of spending is commonly addressed by removing or top-coding high-cost cases. However, this leads to an underestimation of realized savings at the top end of the distribution, therefore, lowering incentives for cost reduction. We show that this trade-off can be weakened by an alternative approach in which cost categories that fall out of the scope of the projects' interventions are excluded from the dependent variable. We find that this approach leads to improvements in precision and model fit that are of the same magnitude as excluding high-cost cases altogether. At the same time, it sharpens the incentives for cost reduction because the model better reflects the costs that projects can affect.

**Keywords** Integrated care · Shared savings · Cost analysis · Population-based financing · Risk adjustment

**JEL Classification** I18 Government Policy · Regulation · Public Health

## Introduction

### Estimating savings of 12 pilot projects on integrated care

In response to the fragmented delivery of health services, the Belgian government launched in 2016 the Plan 'Integrated Care for a better health'. This program supports 12 regional pilot projects during a 5-year period (2018–2022) to initiate a development towards integrated care, i.e. a process of

coordination and delivery of care services so that patients receive seamless and continuous care, tailored to their needs [1]. Participation was voluntary and the 12 pilots have been selected after a nationwide call.

The conceptual model of the national plan is based on the 'Chronic Care Model' [2, 3] and focusses on effective and high quality care for the chronically ill. It is proposed that medical and non-medical care for patients with chronic illnesses is transformed from acute and reactive care to a proactive, planned approach. This should be achieved through effective care provided by a care team and the stimulation of self-management. The model further emphasizes intensive use of patient registries and supporting information technology. After wide consultation of Belgian stakeholders, the model was translated into a policy framework for the care of the chronically ill and a national plan with a series of concrete actions. As other integrated care programs, the actions represent the transformation processes required to achieve a 'Triple Aim', i.e. the improvement of health outcomes and quality of care while being cost-effective or even cost-saving at the same time. The pilot projects serve as test regions to

✉ Karen Geurts  
karen.geurts@intermut.be

Marc Bruijnzeels  
m.bruijnzeels@jvei.nl

Erik Schokkaert  
erik.schokkaert@kuleuven.be

<sup>1</sup> IMA Intermutualistic Agency, Brussels, Belgium

<sup>2</sup> Department of Public Health and Primary Care, Leiden University Medical Centre, The Hague, The Netherlands

<sup>3</sup> Department of Economics, KU Leuven, Leuven, Belgium

identify best practices that can be implemented nationwide in the future.

Each pilot consists of a local consortium that comprises primary, secondary and social care partners, hospitals, local authorities and health insurers operating in the region. Together they cover 21% of the Belgian population. During a preparatory period of 2 years (2016–2017), the selected pilots have designed detailed action plans that address the local health needs and cover a wide range of components of integrated care, such as prevention and patient empowerment, care coordination and continuity, shared patient records and innovative ways of financing. The actions officially went into operation from July 2018 onwards.

Besides a subsidy for management,<sup>1</sup> the pilot projects would receive part of the savings they realize to reinvest in new actions. As in other shared savings programs [4, 5], savings are calculated by comparing average health care spending in the project to a benchmark that is based on predicted levels of spending. For reasons that will become clear later on, the actual payments of savings have been suspended, but the relevant estimations are still calculated as part of the evaluation of the projects. Changes in health care spending are evaluated at the level of the total population living in the pilot region. The calculations are intended to reflect short-term effects, i.e. the impact on spending conditional on time-varying individual characteristics such as health status and socio-demographic features. Prices are fixed at the national level, so savings on spending can only be realized by changes in the utilization of care or by substitution from higher- to lower-cost services, and not by lower prices. Most providers are reimbursed in a fee-for-service system. New ways of financing through bundled payment were planned by some projects but have not yet been implemented. Short-term savings may be achieved through, for example, eliminating medically unnecessary or harmful care ('low value services'), avoiding duplication of services, or pooling scattered financial governance at the regional level (financial integration). Preventive interventions, follow-up care, or patient self-management may also lead to a reduction of avoidable care in the short term.<sup>2</sup> Because these arrangements require new investments, the impact of integrated care on total health spending at population level is unclear. Integrated care is expected to have also long-term cost effects, in particular through improved population health. Such effects,

however, will be less observable in the 5-year evaluation period of the projects.

The Intermutualistic Agency (IMA), a collaborative data and research center set up by all the sickness funds, has been asked by the health authorities to estimate these savings. As in similar studies, the analysis is based on a difference-in-differences approach. Changes in mean spending in the project are compared with risk-adjusted mean changes in a control group, where the latter represent the expected outcome if the project had not been implemented (see, e.g., [6–14]). Using this approach, savings were estimated with large standard errors and were found to be very sensitive to high-cost patients. Therefore, they were considered unsuitable as a basis for actual payments to projects.<sup>3</sup> Part of the explanation of the large standard errors is the strongly skewed distribution of health care costs with a small fraction of the population accounting for a large share of total health spending and mean costs well above median cost. Variance is even more concentrated in the high-cost cases, and exacerbated by the presence of large outliers. As a result, even sophisticated models with large sets of control variables are faced with a portion of high costs not well captured by the model [15]. In this paper we compare different solutions to address this problem.

### Treatment of high-cost patients

A common method to reduce model sensitivity to skewness is the logarithmic transformation of the dependent variable. However, the literature on the comparative performance of models for health care concludes that this transformation is less suitable for estimating individual's health care costs. See [16] for an overview. Using Monte Carlo analyses with hypothetical cost data, or a quasi-Monte Carlo design with empirical data, these studies find that the log regression model generally performs poorly in terms of bias and predictive accuracy, while linear OLS is among the better performing models, in particular for very large sample sizes as the one we use in our study. Moreover, the transformed data provide results on the log scale, while empirical applications as ours typically require results that are expressed in terms of actual costs. New problems arise with the retransformation of costs to the original scale, in particular if there is heteroskedasticity in the data on the transformed scale [16, 17]. Other problems associated with the logarithmic transformation of the dependent variable are the treatment of observations with 0€ costs and the consistent estimation of the confidence intervals. See, e.g., [16, 18]. In line with

<sup>1</sup> The subsidy corresponds to the salary of one full-time employee plus overhead.

<sup>2</sup> Some examples of these interventions are: avoiding duplication of services, e.g. of blood sample or medical images; preventive interventions such as fall prevention in the elderly; follow-up care such as after patients' discharge from a hospital; patient self-management such as medication schedules.

<sup>3</sup> Because the estimations did not yield significant results, payments based on savings have been suspended for the first two years of the program (2018–2019).

these findings, we use a linear model estimated by Ordinary Least Squares. It is the commonly used method in recent evaluations of similar savings programs. See, e.g., [10–14].

Another statistical approach to high-cost observations is to identify them as outliers and reweight them or remove them from the analysis. This indeed leads to more robust estimates but also to an underestimation of potential savings at the top end of the distribution. Integrated care initiatives focus largely on the chronically ill, which are often patients with high levels of spending. Moreover, short-term savings by integrated care are mainly expected from the reduction of unnecessary and avoidable care, which is likely to be concentrated among high-cost patients. Reweighting or removing high-cost patients entails the risk of excluding a significant portion of the costs affected by the initiatives and could thus lead to a systematic underestimation of realized savings.

A less extreme solution is the use of a top-coded dependent variable. Individual expenditures are then limited to a threshold: if a patient's costs exceed a certain threshold, the portion above the threshold does not count. Hence, the top-coded dependent variable is the minimum of actual spending and the threshold. Because of its flexibility and transparency, top-coding has often been used in shared savings models to define spending benchmarks [4, 5, 19]. The thresholds are usually based on the overall cost distribution and are an element of negotiation with the partners of the program. The thresholds can be chosen on the basis of insights in the specific care groups that lead to high spending and as a function of the aims of the program. The treatment of high-cost cases is hence not imposed solely on the basis of statistical considerations, but is motivated by an understanding of the actual scope of the program.

Top-coding has the effect of lowering the impact of high-cost cases, while at the same time retaining these patients in the model. It is preferred to dropping high-cost observations altogether, because at least a portion of the high costs is often predictably associated with specific conditions captured by the covariates. Yet, while less drastic than simply removing observations, top-coding still lowers the incentives for cost control. This implies that we are facing a trade-off here: removing or top-coding high-cost cases may increase robustness of the estimates, but at the same reduce incentives for cost reduction.

We explore another method to weaken this trade-off. To increase reliability and precision of the estimates without excluding a meaningful group of patients from the analysis, we remove from the cost variable those health services that lead to high-cost cases but are unlikely to be affected by the program. The approach bears analogy to the modification of the dependent variable in certain empirical risk-adjustment models [15]. If the program cannot be held responsible for the growth of certain healthcare costs, they should not be part of the dependent variable. The identification of such

cost will depend on the global context of the health insurance system and on the scope of the actual integrated care programs. In our case, we identified two distinct types of health services that entail high spending and are out of the scope of the projects' interventions. Excluding these costs from the dependent variable should sharpen the incentives for cost reduction because the resulting model better reflects the costs that projects can affect.

From a theoretical point of view, increasing reliability and precision of the estimates by tailoring the cost variable to the scope of the program is preferable to removing or top-coding high-cost cases. The aim of this study is to statistically evaluate the different approaches. We compare their performance on measures of precision and model fit, and the stability of the coefficient estimates.

## Method

We use a linear difference-in-differences (DiD) model to compare average changes in spending in the pilot project and the control group. The pre-intervention period ( $t=0$ ) includes observations in 2015, 2016 and 2017, and the postintervention year is 2018 ( $t=1$ ). Observations in the three pre-intervention years are given a weight of 1/3. The model is designed to capture short-term effects and to evaluate whether the pilot project has reduced spending conditional on time-varying health conditions and socio-demographic status.

For each project, we estimate the following linear regression:

$$y_{it} = \beta_0 + \beta_1 P_{it} + \beta_2 T_{it} + \beta_3 (P_{it} \times T_{it}) + \sum_k \delta_k x_{it}^k + \sum_k \gamma_k (x_{it}^k \times T_{it}) + \varepsilon_{it}, \quad (1)$$

where  $y_{it}$  is spending of individual  $i$  in period  $t$ .  $P_{it}$  is a dummy variable which equals one if  $i$  is in the pilot project at time  $t$  and zero otherwise. The coefficient  $\beta_1$  is the pilot fixed effect. It captures group-specific effects, i.e. (non-random) differences between the pilot and the control group in the pre-intervention period.  $T_{it}$  is a dummy variable for the postintervention year.  $\beta_2$  captures time-specific effects, i.e. aggregate factors that cause changes in  $y_{it}$  regardless of the group to which the individual belongs. The interaction term ( $P_{it} \times T_{it}$ ) is equal to one only for individuals in the pilot in the postintervention year. The coefficient of interest,  $\beta_3$ , is an estimate of the effect of the pilot project on changes in spending between the pre- and postintervention periods. We estimate the DiD model with OLS, the common method for the estimation of risk-adjusted health care costs, especially for large datasets (see also [10–14]). This setting allows us to use the maximum number of control observations, to

flexibly include control variables and different time periods, and to obtain direct estimates of the effect of the projects on spending.

The DiD model makes the identifying assumption that in the absence of treatment, the average outcomes for the treated and control groups would have followed parallel trends over time [20, 21]. This requires us to define a set of covariates,  $x_{it}^k$  ( $k = 1, \dots, n$ ), that affect spending and are related to the intervention. With the inclusion of these covariates, the model should control for a sufficient proportion of variation in the data. We include a rich set of individual-level variables that risk-adjustment models have shown to be good predictors of total healthcare spending and which control, at the same time, for differences in the composition of the population of the pilot and control groups. The set of covariates is inspired by the Belgian risk-adjustment model for the financing of the health insurers [22]. In addition to interactions of age classes and sex, we include socio-demographic and socioeconomic variables, geographical variables, a dummy for mortality, a set of covariates indicating care dependence, and a large number of morbidity indicators. Care dependence and morbidity indicators capture chronic conditions. Care dependence variables are based on insurance entitlements related to disability, long-term nursing care, physiotherapy and revalidation, and morbidity indicators are pseudo-diagnoses based on ambulatory and inpatient drug prescriptions, some in combination with use of care and facilities for specific conditions. The full list of variables is provided in Appendix 1. Finally, we include municipality-level fixed effects to control for urbanization and medical supply in the area.

As a validity check of the main model assumption, we compare changes in spending in the pilot and risk-adjusted control groups in the three pre-intervention years. We also include interactions between the covariates and the time dummy, ( $x_{it}^k \times T_{it}$ ) and keep only the significant ones using backward elimination. The time-interactions allow the effect of a covariate to differ between the pre- and postintervention periods by an amount that is different from the aggregate time effect,  $\beta_2$ . The error term  $\varepsilon_{it}$  represents random variation in spending.

Our baseline model has total (uncorrected) health care expenditures as the dependent variable and makes use of all the available observations. We then compare the results of removing high-cost cases and of excluding some specific cost categories from the dependent variable. Finally, we check the gains obtained by top-coding the dependent variable.

## Data

We use claims data from 2015 through 2018 of all residents covered by the Belgian mandatory health insurance. The data are collected by the health insurers and compiled by the Intermutualistic Agency in a permanent database for research. It includes all reimbursed services as well as patient socio-demographic characteristics and social security related data. The pilot samples consist of all individuals living in the region of the pilot project, while the control sample is the same for all projects and includes all individuals not living in any of the pilot regions. Because we are interested in changes in mean spending across all regional levels of care, no patients are excluded, and individuals are assigned to a project or control group on a year-by-year basis.<sup>4</sup> The project-level estimates are thus based on an unbalanced panel of about 8.9 million observations per year.

The dependent variable is total spending reimbursed under the compulsory health insurance at the individual-year level. Reimbursements represent about 92.5% of total spending in the Belgian compulsory health insurance system, while out-of-pocket payments represent 7.5%.<sup>5</sup>

Table 1 shows summary statistics of the study sample in the pre-intervention period (2015–2017). The number of individuals in a pilot project varies between 96,000 and 362,000, and is 193,000 on average. The control group is 45 times as large and includes more than 8.7 million individuals. Mean spending per project in the pre-intervention period was 2037€ on average across projects, and increased by 6% in 2018. Both measures differ considerably between projects, with projects substantially below and others above the corresponding figures in the control group. The distribution of spending is strongly right skewed. Mean spending is almost 4 times the median. Moreover, mean growth is considerably above median growth, suggesting that increases in spending are disproportionately situated in higher cost classes. The top

<sup>4</sup> The downside of this approach is that individuals who move to another region, transfer previous incentives that might affect health care use to another group. Another approach used in some evaluations is to limit the sample to the cohort that can be assigned to a particular group over the entire period of analysis. However, over an 8-year pre- and postintervention period, this would lead to a high percentage of exclusions, including an important target group of integrated care interventions, in particular older persons who die before the end of the evaluation period.

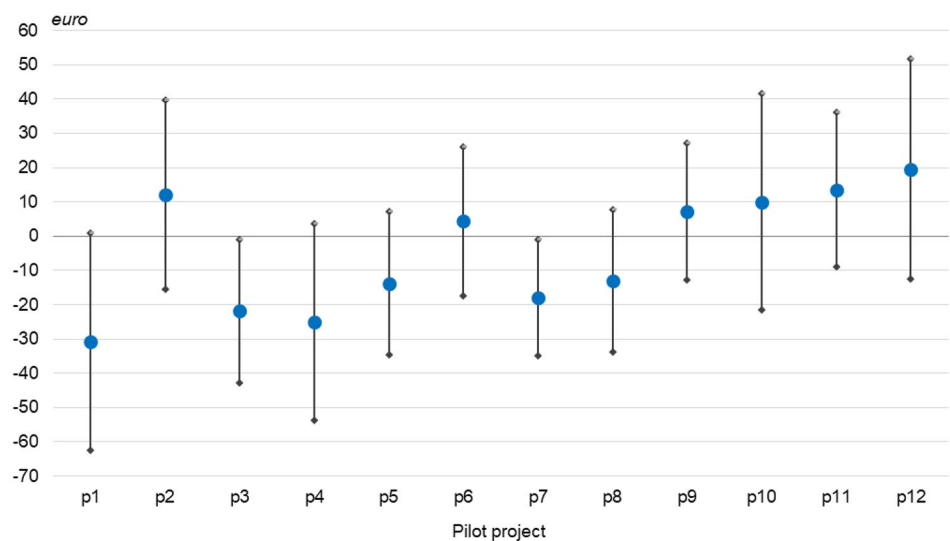
<sup>5</sup> In addition to reimbursed and out-of-pocket payments, there are also supplements. Supplements are the difference between the actual rate and the conventional rate for services included in the compulsory health insurance system, or for the full price for care services not included in that system. Supplements in hospitals are regulated to a certain extent, but in the ambulatory sector providers have a large degree of freedom to determine them.

**Table 1** Summary statistics

	Pilot projects Average across projects (min–max)	Control group
Total number of patients	193,000 (96,000–362,000)	8,709,000
Mean spending per patient		
Pre-intervention period	2037€ (1670€–2361€)	1922€
Growth pre- to postintervention period	+ 6.0% (+ 2.9 to + 7.4%)	6.5%
Median spending per patient		
Pre-intervention period	546€ (443€–622€)	517€
Growth pre- to postintervention period	+ 4.4% (+ 1.8 to + 6.3%)	4.3%

Values across the 12 individual pilot projects and for the control group  
The pre-intervention period is 2015–2017; the postintervention period is 2018

**Fig. 1** The effect of the pilot projects on spending. Baseline model before data preprocessing. Coefficient estimates and 95% confidence intervals of  $\beta_3$



percentile accounts for 19% of total spending and the top 5 percentile accounts for 50% of total spending (see Table 3).

## Results

### Baseline model

We evaluate the effect of the pilot projects on spending using a difference-in-differences model estimated with OLS, as described in the Method section. We first show baseline results without preprocessing the data, that is, no cost groups are excluded from the dependent variable. Figure 1 shows the coefficient estimates and 95% confidence intervals of  $\beta_3$  for each pilot-level model. The results range between cost reductions of 30€ and increases in spending of 20€. Coefficient estimates are not significantly different from zero for ten projects and significantly negative for the two others. The lack of significant results may be due to the early stage

of the projects or to the fact that the  $\beta_3$  coefficients are estimated with relatively large standard errors of 12.5 on average across the 12 project-level estimations. The width of the confidence intervals ranges from 34€ to 64€, or 1.5 to 3.8% of mean spending.

### Excluding the high-cost cases

A traditional solution to increase precision is to exclude high-cost cases from the sample. Specifications 1.a and 1.b in Table 2 show the impact on the estimates when the top 0.1% and the top 1% high-cost cases are excluded, respectively. The table reports two fit statistics to evaluate model performance: the adjusted *R*-square as the fraction of the total variance in the dependent variable explained by the model, and the mean absolute error as the average of the absolute errors.<sup>6</sup> Because we are interested in precise and stable estimates, we also report the standard errors of the

<sup>6</sup> The mean absolute error avoids the disadvantage of *R*-square type measures that outliers are heavily weighted.



**Table 2** Model performance before data preprocessing

Dependent variable: all cost groups included	Model fit statistics		Coefficient $\beta_3$	
	Adjusted R-squared	Mean absolute error	Standard error	Absolute coefficient change w.r.t. Baseline
Specification 1—baseline	0.31	1544	12.5	
Specification 1.a—top 0.1% high-cost cases excluded	0.45	1424	8.5	€6 (1–16)
Specification 1.b—top 1% high-cost cases excluded	0.47	1167	6.0	€14 (1–32)

Average values across the 12 project-level estimations with minimum and maximum values in parentheses

estimated coefficient,  $\beta_3$ , and the absolute parameter change compared to the baseline specification. Each cell provides the average calculated across the 12 project-level estimations. For the last measure, we also report the minimum and maximum value across the projects.

After excluding the top 0.1% and the top 1% high-cost cases, model fit measures strongly improve and standard errors are reduced on average by 32 and 52%. The last column, however, shows that the results are highly sensitive to these high-cost cases: parameter estimates substantially fluctuate across the specifications, at least for part of the projects.

### Spending by care type among high-cost patients

To understand which cost groups cause unexplained variation at the top end of the distribution, we analyze the distribution of spending by type of care. The aim is to identify those cost groups that lead to high-cost patients, but which

might be out of the scope of the projects. If the projects cannot be held responsible for the growth of certain costs, these costs should not be part of the dependent variable. Table 3 shows a summary of our findings. It reports the distribution of spending by care type in patients with lower costs (the 1st–94th percentile) and in the 5 highest percentiles. The latter account for 50% of total spending. Three groups of care services stand out in the spending by high-cost patients: Medication, Nursing home stays and, to a lesser extent, Nursing care and physiotherapy (services outside nursing homes). Together, these services represent 56–79% of spending in higher cost groups, but only 36% among low-cost patients.

Nursing home stays and nursing care are at the core of the transformation of health care towards integrated care. These services play a crucial role both in horizontal initiatives of the pilot projects (such as care coordination after hospitalizations, multidisciplinary care of the chronically ill, or reinforcement of primary care to avoid unnecessary

**Table 3** Distribution of spending by care type among low-cost (1st–94th percentiles) and high-cost patients (95th–99th percentiles)

	Percentiles of total spending					
	0–94th	High-cost patients				
		95th	96th	97th	98th	99th
Share in total spending (%)	50.4	5.0	6.4	8.3	10.7	19.2
Total	100	100	100	100	100	100
By care type						
Consultations	17	6	4	4	3	2
Fees of medical specialists	30	24	20	15	10	12
Clinical biology and medical imaging	16	11	9	7	5	6
Nursing care and physiotherapy	14	22	20	20	14	14
Nursing home stays	1	5	11	27	49	20
Medication and medical devices	22	29	32	24	16	44
Medication outside the hospital lump sum	2	12	18	12	7	24
Life-saving medical devices	3	7	5	4	3	12
Other	17	11	9	8	6	8
Other	1	3	3	3	2	2

Total sample (national population), 2016

**Table 4** Model performance after data preprocessing

Dependent variable: two subsets <sup>a</sup> of medication excluded from costs	Model fit statistics		Coefficient $\beta_3$	
	Adjusted <i>R</i> -squared	Mean absolute error	Standard error	Absolute coefficient change w.r.t. baseline
Specification 2—baseline	0.46	1231	7.3	€10 (0–31) <sup>b</sup>
Specification 2.a—costs top-coded at 99.9th percentile	0.51	1212	6.7	€2 (0–7)
Specification 2.b—costs top-coded at 99th percentile	0.52	1150	6.0	€4 (0–8)

Average values across the 12 project-level estimations with minimum and maximum values in parentheses

<sup>a</sup>Subsets of medication: medication outside the hospital lump sum and life-saving medical devices

<sup>b</sup>Absolute coefficient change with respect to specification 1—baseline (see Table 2)

complications), as well as in dedicated actions for dependent patients (such as prevention coaches or care trajectories to allow older people to live longer at home). Excluding these services would clearly ignore an important part of the effect of the pilots on spending.

The link between the use of ‘medication and medical devices’ and the integrated care actions is more ambiguous. On the one hand, several initiatives of the projects focus on medication compliance and avoidance of medication overuse. On the other hand, treatments for specific conditions, often with new and expensive medicines, clearly fall outside the scope of the projects in the short term. We identified two subsets that the integrated care projects are unlikely to affect. The first is ‘Medication outside the hospital lump sum’. It concerns a list of ATC-codes for which the Belgian health authorities have laid down that the cost should not reduce their use. They are considered to be “medicinal products whose active substance is of major importance in medical practice, taking into account therapeutic and social needs and the innovative nature of the ingredient”. These medicines are kept outside the lump sum that patients pay for hospital drugs because the high cost could otherwise strongly inhibit their prescription.<sup>7</sup> Spending on these medicines has been increasing very sharply in recent years. Between 2016 and 2018, the increase was 20%, compared to a 6% growth of total spending. The pilot projects certainly cannot be held responsible for disparities in the growth of these costs, it even would run counter the official health care policy if the projects would reduce the use of these medicines. Keeping these

costs in the sample entails the risk that changes in spending are wrongly attributed to the effect of the program. A same reasoning can be followed for a second subset of medication that we labeled as life-saving medical devices. It includes implants, implantable cardiac defibrillators, human body material, and dialyses. Even if the projects invest in prevention and avoidance of complications that could reduce the need of this material in the long term, they are unlikely to affect their use in the short term. Both medication outside the hospital lump sum and life-saving medical devices represent a disproportional share of spending among high-cost patients, i.e. 10–36%, while only 5% among low-cost patients. We exclude these cost subsets from the dependent variable, as they are problematic from both a statistical and context-specific viewpoint.

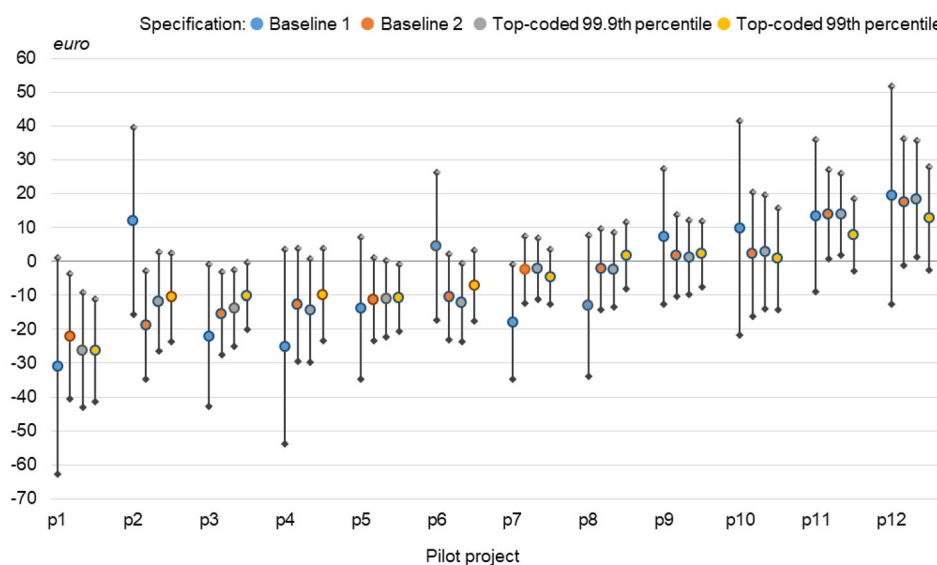
### Redefining the dependent variable

The results of redefining the cost variable are shown in Table 4. In specification 2, we do not exclude any patients but we remove the two subsets of medication discussed above, i.e. medication outside the hospital lump sum and life-saving medical devices, from the dependent variable. The model fit measures strongly improve compared to specification 1—baseline, and the standard errors are reduced by 42%. Remarkably, the improvements are of the same magnitude as in specification 1.b, where the top 1% patients are excluded (see Table 2). This indicates that the excluded cost groups are indeed responsible for a large part of the unexplained variation at the top end of the distribution.

Figure 2 illustrates the estimates and the standard errors for the  $\beta_3$  coefficients of the 12 projects. The point estimates of the coefficients in specification 2 substantially differ from specification 1. Although standard errors are greatly

<sup>7</sup> As an instrument to control hospital drugs costs, medication during hospital stays is reimbursed on the basis of a fixed payment per patient-day (per diem). However, a number of mostly expensive and/or innovative drugs are reimbursed on a fee-for-service basis. These drugs are defined by their active substance (5th level of ATC classification).

**Fig. 2** Comparison of different specifications of the model. Specifications as in Tables 2 and 4. Coefficients and 95% confidence intervals of  $\beta_3$



reduced, they remain relatively large. The width of the confidence intervals still ranges from 19€ to 37€, or 0.9–2.6% of mean spending. We will now check whether top-coding can remedy this problem.

### The additional gains from top-coding

In specifications 2.a and 2.b (see Table 4), the dependent variable is top-coded at the value of the 99.9th and 99th percentiles of the cost distribution, respectively.<sup>8</sup> If a patient's cost exceeds this threshold, it replaced by the threshold. The last two rows of Table 4 show that top-coding moderately improves model fit measures and further reduces standard errors by 8 and 11%, respectively. More importantly, we find that coefficient estimates remain fairly stable. They differ from specification 2 results by only 2€–4€ on average, with a maximum of 8€. Figure 2 provides a visual representation of this stability. Since top-coding affects the incentives for cost reduction at the highest end of the distribution, it could be considered not to implement it.

### Validity check: the parallel trends assumption

As a validity check of the parallel trend assumption, we compare pre-intervention trends between the pilot and control groups. The graphs in Appendix 2 (Fig. 3) show mean spending in the pilot projects and risk-adjusted spending in the control groups from 2015 to 2018. Although the same control group is used for all projects, risk-adjusted mean

spending differs for each pilot-level estimation, both in the level of spending and spending growth. The included risk-adjusters adequately control for differences in the composition of the populations of the pilot and control groups: the graphs show that pre-intervention trends in both groups are very similar. Moreover, the results for 2018 show no clear deviation from these parallel trends. Indeed, even if standard errors are strongly reduced after redefining the dependent variable, we still find that the estimated effects of the pilot projects on spending are not or only just significantly different from zero (see Fig. 2).

## Discussion

In principle the exclusion from the dependent variable of expensive cost items that cannot be influenced by the integrated care projects does not only improve the statistical properties of the estimation exercise but is also justifiable on conceptual grounds, as it can sharpen the incentives to control costs. The identification of the costs to be excluded is not an easy task, however. Integrated care models are complex interventions and the mechanisms underlying the development of costs are difficult to trace. Excluded costs, e.g. high-cost medicines, will almost never be fully independent from the services of integrated care models. If the interdependence is substantial, the exclusion of expensive drugs could also reduce the incentives to control costs. The trade-off between

<sup>8</sup> The values of the 99.9th and 99th percentile are about 19 and 11 times the mean cost.



statistical robustness and strong incentives will not fully disappear in this approach.

It is therefore obvious that the decision about which costs to exclude needs solid justification. For the reasons explained earlier, we believe that in our specific exercise we can safely assume that the indirect effects are not very large, especially in the short-term frame of our evaluation. Yet, our approach should be tested and further refined as more data on the results become available.

The results based on preprocessed data suggest that one project realized significant savings, while for the others, estimates are not significant or only just different from zero. It remains unclear whether these results are acceptable as a reliable basis for payments to projects. The lack of clear significant results may be due to the early stage of the program. Savings in similar programs have been found to increase over time [10–12, 14]. Moreover, because prices for care services in Belgium are fixed, savings on spending can only be realized by lower utilization of services and by substitution from higher- to lower-cost services, which is likely to take more time than savings by lower prices. The short time period is one of the main limitations of this study.

Consequently, one reason for the small effects we observe may be that, in a transition period towards integrated care, realized savings and increased costs may coexist, for example by strengthening primary care and multidisciplinary collaboration to ensure better patient follow-up or avoid duplication of services. An interesting way to move forward would be to investigate specific groups of care services that are mostly affected in either direction in this initial stage and to explicate their impact on the total cost development. DiD findings such as those shown in Fig. 3 offer a natural starting point for such a deeper analysis of the differences in the cost developments between the projects compared and the risk-adjusted control group. At this stage, our results are not sufficiently robust for a reliable analysis of this kind, but we plan to focus on the decomposition of costs in later work.

Finally, more effort should be invested in collecting information on quality of care and health outcomes. The quality and health data will be less skewed so that estimates of the program effects are likely to be more robust [10–12]. This will make it possible to analyze whether the integrated care projects have successfully achieved the intended Triple Aim of improving quality and health while reducing costs.

## Conclusion

Calculations of the effect of care programs on spending that are used to reward organizations which are found to realize savings, place high demands on the precision of the estimates. Relatively small standard errors in statistical terms often represent substantial budgets when expressed in monetary values. It may lead to undue payments to less-performing organizations and undercompensation of others. The calculation of savings and losses in the Belgian integrated care program is, as in shared savings programs, based on comparisons of mean spending growth in the program and a control group. These means are highly sensitive to high-cost patients. A purely statistical approach, which reweights high-cost cases or removes them from the analysis may lead to a systematic underestimation of realized savings because short-term cost effects of integrated care are likely to be achieved in patients at the top end of the distribution. Our proposed approach combines statistical treatment of high-cost cases with careful pre-processing of the data, incorporating insights in the health care costs that integrated care actions in the specific context of the program are likely to affect. Care services that the program cannot be held responsible for are removed from the dependent variable. Removal of these services strongly improves model fit measures and leads to more precise estimates of savings and losses without excluding a meaningful group of patients from the analysis. We further check the sensitivity of the results to high-cost cases by re-estimating the model with a dependent variable that is top-coded at some level. Top-coding further increases the robustness of the estimated coefficients, but the additional gains are modest. Moreover, the results are fairly stable to top-coding. Since top-coding reduces the incentives for reducing costs and for focusing on the chronically ill patients with high expenditures, it may be considered to refrain from it.

We find that defining a dependent variable that is tailored to the context of the program yields goodness-of-fit measures that are similar to simply excluding the top 1% cost cases from the sample. On the other hand, changes in spending are still estimated with relatively large standard errors. The width of the confidence intervals ranges between 0.9 and 2.6% of mean spending. These may be relatively precise results from an economic evaluation perspective, but they correspond to large amounts of money at project-level. Even if significant savings were estimated in subsequent years, random factors would lead to substantial differences in actual payments.

## Appendix 1: List of covariates

Socio-demographic and -economic variables.

- Sex-age categories (0; 1–25; 26–50; 51–75; 75+ year)
- Deceased in year  $t$
- Single
- Self-employed
- Entitled to increased compensation
- Dependent on subsistence level income or social welfare

Care dependence variables (based on insurance entitlements).

- Unable to work (occupationally disabled)
- Disabled
- Disabled and care dependent
- Heavily care-dependent patient receiving long-term nursing care at home (3 levels of dependency)
- Receiving physiotherapy for severe pathology
- Receiving rehabilitation therapy

Morbidity indicators based on ambulatory and inpatient drug prescriptions.

- Cardiovascular disease—general
- Cardiovascular disease—heart disease
- Thrombosis—anticoagulants
- Chronic obstructive pulmonary disease (COPD)
- Asthma
- Diabetes with cardiovascular disease

- Psoriasis
- Crohn's disease, Colitis Ulcerosa, Psoriatic arthritis, Rheumatoid arthritis
- Psychoses in persons of 70 years and less
- Psychoses in persons over 70 years of age
- Parkinson's disease
- Epilepsy and neuropathic pain
- HIV
- Multiple sclerosis
- Alzheimer's disease
- Hepatitis B or C
- Thyroid disorders
- Exocrine pancreatic disease

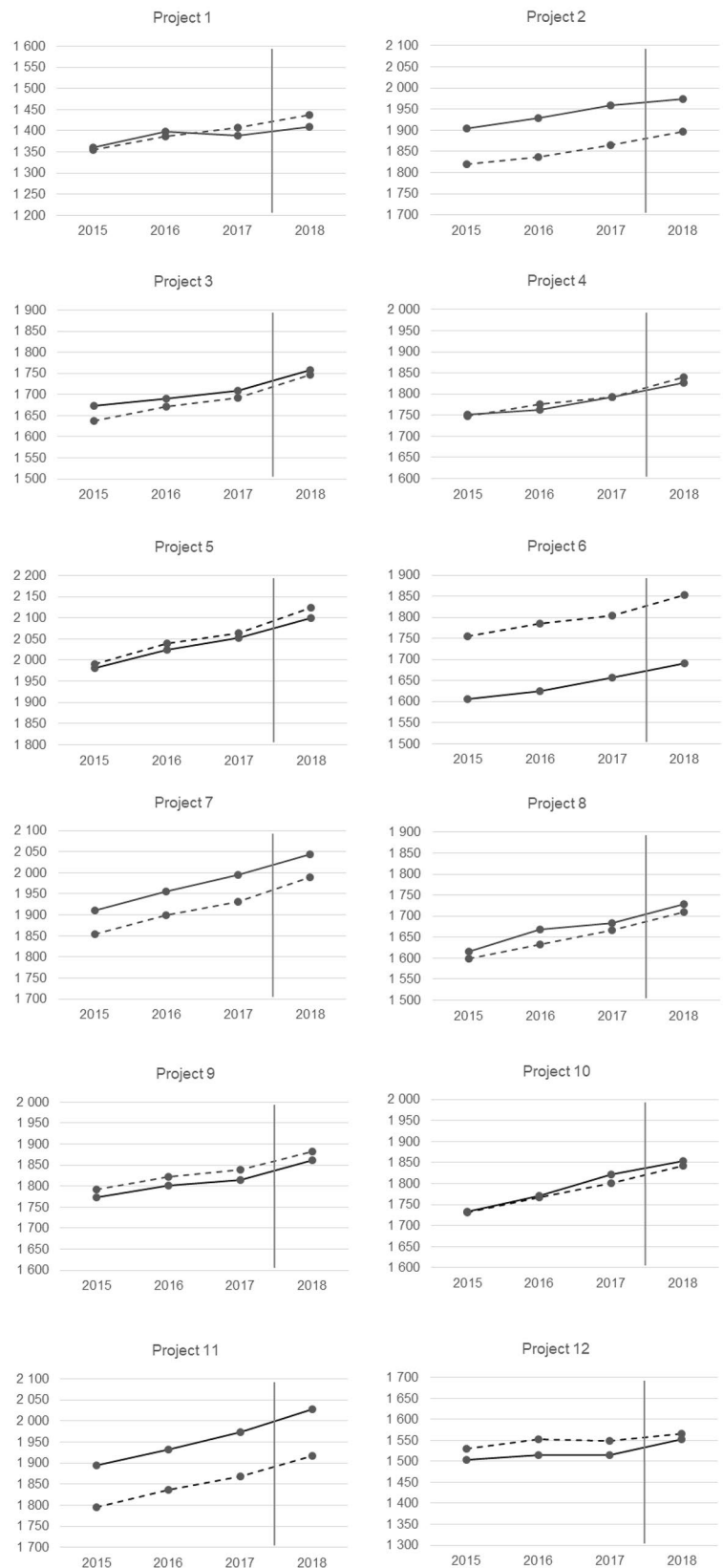
Morbidity indicators based on ambulatory and inpatient drug prescriptions and on use of care and facilities for specific conditions.

- Diabetes mellitus with insulin
- Kidney failure
- Cancer
- Hemophilia
- Cystic fibrosis
- MS/ALS/Huntington

Geographical variables.

- Region (Brussels, Flanders, Wallonia)
- Level of urbanization of municipality
- Level of medical supply of municipality

**Fig. 3** Mean spending (in euro) in the pilot projects (solid lines) and in the risk-adjusted control groups (dotted lines). Dependent variable as in specification 2—baseline (see Table 4: two subsets of medication are excluded)



## Appendix 2: Parallel pretrends

See Fig. 3.

**Acknowledgements** This paper reports results that are obtained within a project between the Belgian Health Authorities and IMA (Intermutualistic Agency). The authors alone are responsible for its content. We thank the participants of the 2020 RAN-workshop, in particular Randy Ellis and Tom McGuire, for valuable comments.

## References

- De Ridder, R., Bourgeois, J., Van den Bogaert, S., Van der Brempt, I.: Implementing integrated care in Belgium: a nationwide mobilization. *Int. J. Integr. Care* (2017). <https://doi.org/10.5334/ijic.3441>
- Wagner, E., Austin, B., Von Korff, M.: Organizing care for patients with chronic illness. *Milbank Q.* **74**(4), 511–544 (1996)
- Paulus, D., Van den Heede, K., Mertens R.: Position paper: organisatie van zorg voor chronisch zieken in België. Brussels, KCE Report 190A (2012)
- Bailit, M., Hughes, C.: Key design elements of shared-savings payment arrangements. The commonwealth fund issue brief. [http://www.bailit-health.com/publications/082111\\_bhp\\_key\\_designelements\\_sharesavings.pdf](http://www.bailit-health.com/publications/082111_bhp_key_designelements_sharesavings.pdf) (2011). Accessed 2 Apr 2021
- Medicare Shared Savings Program: Shared savings and losses and assignment methodology. <https://www.cms.gov/files/document/shared-savings-losses-assignment-spec-v8.pdf-0> (2020). Accessed 2 Apr 2021
- Vats, S., Ash, A., Ellis, R.: Bending the cost curve? Results from a comprehensive primary care payment pilot. *Med. Care* **51**(11), 964–969 (2013). <https://doi.org/10.1097/MLR.0b013e3182a97bdc>
- Nolte, E., Pitchforth, E.: What is the evidence on the economic impacts of integrated care?. Technical Report. World Health Organization, Copenhagen, Denmark. <https://researchonline.lshtm.ac.uk/id/eprint/2530944> (2014). Accessed 2 Apr 2021
- Busse, R., Stahl, J.: Integrated care experiences and outcomes in Germany, the Netherlands, and England. *Health Aff.* (2014). <https://doi.org/10.1377/hlthaff.2014.0419>
- Peckham, A., Rudoler, D., Allin, S., Bhatia, D., Abdelhalim, R., Kavelaars, R.A., Marchildon, G.: Accountable care organizations: success factors, provider perspectives and an appraisal of the evidence. North American Observatory on Health Systems and Policies, Toronto. Rapid review (no. 12) (2019)
- Song, Z., Safran, D.G., Landon, B.E., Landrum, M.B., He, Y., Mechanic, R.E., Day, M.P., Chernew, M.E.: The ‘Alternative Quality Contract’, based on a global budget, lowered medical spending and improved quality. *Health Aff.* **31**(8), 1885–1894 (2012). <https://doi.org/10.1377/hlthaff.2012.0327>
- Song, Z., Rose, S., Safran, D.G., Landon, B.E., Day, M.P., Chernew, M.E.: Changes in health care spending and quality 4 years into global payment. *N. Engl. J. Med.* **371**, 1704–1714 (2014). <https://doi.org/10.1056/NEJMsa1404026>
- Song, Z., Ji, Y., Safran, D.G., Chernew, M.E.: Health care spending, utilization, and quality 8 years into global payment. *N. Engl. J. Med.* **381**, 252–263 (2019). <https://doi.org/10.1056/NEJMsa1813621>
- McWilliams, J.M., Landon, B.E., Chernew, M.E.: Changes in health care spending and quality for Medicare beneficiaries associated with a commercial ACO contract. *JAMA* **310**(8), 829–836 (2013). <https://doi.org/10.1001/jama.2013.276302>
- McWilliams, J.M., Hatfield, L., Landon, B.E., Hamed, P., Chernew, M.E.: Medicare spending after 3 years of the Medicare Shared Savings Program. *N. Engl. J. Med.* **379**, 1139–1149 (2018). <https://doi.org/10.1056/NEJMsa1803388>
- McGuire, T.G., van Kleef, R.C.: Risk sharing. In: McGuire, T.G., van Kleef, R.C. (eds.) *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, pp. 105–132. Academic Press, Cambridge (2018)
- Jones, A.M.: Models for health care. In: Clement, M.P., Hendry, D.F. (eds.) *The Oxford Handbook of Economic Forecasting*. Oxford University Press, Oxford (2011)
- Ellis, R.P., Martins, B., Sherri, R.: Risk adjustment for health plan payment. In: McGuire, T.G., van Kleef, R.C. (eds.) *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, pp. 55–104. Academic Press, Cambridge (2018)
- Manning, W., Mullahy, J.: Estimating log models: to transform or not to transform? *J. Health Econ.* **20**, 461–494 (2001). [https://doi.org/10.1016/S0167-6296\(01\)00086-8](https://doi.org/10.1016/S0167-6296(01)00086-8)
- Hayen, A.P., van den Berg, M.J., Meijboom, B.R., Struijs, J.N., Westert, G.P.: Incorporating shared savings programs into primary care: from theory to practice. *BMC Health Serv. Res.* (2015). <https://doi.org/10.1186/s12913-015-1250-0>
- Angrist, J.D., Pischke, J.S.: *Mostly Harmless Econometrics. An Empiricist’s Companion*. Princeton University Press, Princeton (2009)
- Wooldridge, J.: *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge (2001)
- Schokkaert, E., Guillaume, J., van de Voorde, C.: Risk adjustment in Belgium: why and how to introduce socioeconomic variables in health plan payment. In: McGuire, T.G., van Kleef, R.C. (eds.) *Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice*, pp. 209–232. Academic Press, Cambridge (2018)

**Publisher’s Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.