Erasmus School of
Health Policy
& Management

Implications of COVID-19 for Risk Adjustment of Health Plan Payment

December 2020

Richard C. van Kleef
Mieke Reuser

Erasmus University Rotterdam

Implications of COVID-19 for Risk Adjustment of Health Plan Payment

Richard C. van Kleef [1]

Mieke Reuser [1,2]

[1] *Erasmus School of Health Policy & Management (Rotterdam, The Netherlands)*

[2] *Dutch Ministry of Health, Welfare & Sports (The Hague, The Netherlands)*

Corresponding author: Richard C. van Kleef (vankleef@eshpm.eur.nl)

ESHPM working paper: December 10, 2020

Abstract: The COVID-19 pandemic has led to disruptions in healthcare utilization and spending. While some changes might persist (e.g. substitution of specialist visits by online consultations) others will be transitory (e.g. fewer surgical procedures due to postponement or cancellation of treatments). This paper discusses the implications of transitory changes in healthcare utilization and spending for risk adjustment of health plan payment. In practice, risk adjustment methodologies typically consist of two steps: 1) calibration of payment weights for a given set of risk adjusters and 2) calculation of payments to insurers by combining the calibrated weights with enrollee characteristics. In this paper we first introduce a simple conceptual framework for analyzing the (potential) distortions from COVID-19 for both steps and then provide a hypothetical illustration of how these distortions can lead to under- or overpayment of insurers. The size of these under/overpayments depends on 1) the impact of COVID-19 on patterns in utilization and spending, 2) features of the risk adjustment system, 3) the distribution of risk types across plans, and 4) the extent to which insurers are disproportionately affected by COVID-19. In a final step, we qualitatively discuss how COVID-19 distorts risk adjustment in the Netherlands and how a series of (temporary) policy measures can be used to mitigate the (effects of this) distortion.

Acknowledgements: The authors are grateful to CSS Insurance for providing administrative claims data and to Lukas Kauer for preparing this data. The authors thank Konstantin Beck, Randy Ellis, Lukas Kauer, Tom McGuire, Sonja Schillo, Wynand van de Ven, René van Vliet and participants of the online meeting of the Risk Adjustment Network 2020 for their valuable comments on earlier versions of this paper.

1. Introduction

The COVID-19 pandemic has led to disruptions in healthcare utilization and spending. During the pandemic, utilization of services related to treatment of COVID-19 such as intensive care and specific medical equipment has been higher than normal. On the other hand, utilization of regular services such as surgical procedures has been lower due to postponement, cancellation or substitution of medical treatments. While some of these changes might persist others are likely to be transitory. This paper focusses on the latter and shows how transitory disruptions in healthcare utilization and spending, even if short-lived, can have implications for risk adjustment of health plan payment for years to come.

Risk adjustment systems provide insurers with payments based on risk characteristics of their enrollees called ‘risk adjusters’. Common risk adjusters are age, gender and morbidity flags based on (prior) diagnoses, procedures or spending. Each of these risk adjusters comes with a ‘payment weight’ or ‘risk score’, which is typically higher for the elderly and the chronically ill than for the young and healthy. In practice, risk adjustment procedures roughly consist of two steps: 1) calibration of payment weights for a given set of risk adjusters and 2) calculation of payments to insurers by combining the calibrated weights with enrollee characteristics. This widely accepted approach to risk adjustment and plan payment depends on an assumption that the relationship between the risk adjuster variables and individual spending is reasonably stable between the time weights are calibrated and the time payments are executed. This paper will show how the COVID-19 pandemic disrupts relationships between risk adjusters and spending and thereby complicates both of the aforementioned steps.

Calibration of payment weights (step 1) is generally done by a least-squares regression of healthcare spending on a predefined set of risk adjusters [1]. In most systems, the dataset used for calibration is different from the population and/or year for which risk-adjusted payments to insurers are calculated. For example, the Netherlands use nationwide spending data from year t-3 (in combination with morbidity flags from t-4 as well as other risk adjusters) to calibrate weights for payment year t [2]. Switzerland uses somewhat more recent spending data, from t-1 in combination with morbidity flags from t-2 as well as some other risk adjusters [3]. Systems in the United States, such as the Marketplaces (Affordable Care Act) and Medicare Advantage rely on calibration data from both a prior period and a subpopulation [4, 5]. Before running the regression, researchers typically modify the calibration data in order to make it representative for the payment year. Typical modifications include corrections for cost inflation and changes in the benefits package and composition of the population. Although there is no doubt that COVID-19 changes patterns in utilization and spending, it is uncertain what these patterns will eventually look like and how they will differ from those in the past and the future. Because of this uncertainty it will be harder to make calibration samples representative, both during the pandemic (when the calibration sample comes from the pre-pandemic period) and after the pandemic (when the calibration sample comes from the pandemic period). We will discuss the circumstances under which discrepancies between the calibration sample and the payment year can be problematic.

Calculation of payments to insurers (step 2) is generally done by combining the payment weights from step 1 with characteristics of health plan enrollees. For each enrollee, the combination of risk adjuster flags and payment weights results in a value for 'predicted spending'. The sum of predicted spending over all individuals enrolled in a plan forms the basis for the payment to the insurer. During the pandemic, the postponement, cancellation and/or substitution of care potentially influence patterns in risk adjuster flags. For example, hospitals might have performed fewer surgical procedures (e.g. because of a temporary shortage of intensive care beds or to mitigate the risk of COVID-19 infections among vulnerable patients), resulting in fewer inpatient diagnostic flags. In prospective risk adjustment systems, morbidity flags are based on data from prior years. For a given set of payment weights, fewer procedures in year t-1 (e.g. 2020) might lead to lower plan payments in year t (2021). We will discuss and illustrate the circumstances under which such a disruption of risk adjuster flags can be problematic.

In sum, the goal of this paper is to identify and discuss the complications of COVID-19 for risk adjustment of health plan payment. Contrary to many existing papers on the design and evaluation of risk adjustment systems for health plan payment (e.g. [6-10]) our focus will be on under/overpayment at the level of *insurers*, rather than the level of individual consumers or specific subgroups of the population. Individual- and/or group-level under/compensation are typically used to quantify incentives for insurers to engage in risk selection, i.e. actions to target (deter) consumers that are predictably (un)profitable. Risk selection by insurers, however, is not our primary interest here. Instead, we analyze the circumstances under which COVID-19-related distortions in plan payments can lead to overpayment or underpayment of individual insurers and/or the market as a whole. The underlying idea of our analysis is that over/underpayment at these levels are problematic given that – in most healthcare systems – the impact of COVID-19 on healthcare spending is considered a 'joint' responsibility of the regulator, insurers and healthcare providers rather than the responsibility of (individual) insurers alone.

The structure of this paper is as follows. Section 2 presents a simple conceptual framework for analyzing how COVID-19 distorts the calibration of payment weights (step 1) and the calculation of payments to insurers (step 2). Section 3 provides a hypothetical quantitative illustration. Section 4 discusses the potential impact of COVID-19 on a real-world risk adjustment system: the one used for health plan payment in the Netherlands. Our analysis shows that COVID-19 is expected to distort the Dutch risk adjustment system in multiple ways and for many years. Section 5 concludes by summarizing our most important findings, discussing the international relevance of these findings, and describing a series of policy measures that could be considered to mitigate the effect of distortions from COVID-19 including (temporary) redesign of current risk adjustment methodology and implementation of risk sharing.

2. Conceptual framework

This section introduces a conceptual framework for analyzing the impact of COVID-19 on risk adjustment. More specifically, we will describe how the pandemic potentially disrupts patterns of healthcare utilization and spending (Section 2.1) and then discuss the implications of disruptions for the calibration of payment weights (2.2) and the calculation of payments to insurers (2.3).

2.1. How COVID-19 can affect utilization and spending

The (potential) effects of the pandemic on utilization and spending are multidimensional. First, during the pandemic, utilization of services directly related to treatment of COVID-19 such as intensive care and protective medical equipment are likely to be higher than before and after the pandemic. Second, during high COVID-19 incidence rates, utilization of 'regular' services such as inpatient surgery might be lower, at least temporarily, due to lock-down measures and/or allocation of existing hospital beds to COVID-19 patients. Third, in periods with decreased or stabilized COVID-19 incidence rates, utilization of regular treatments might be higher than normal because healthcare providers catch up some of the postponed treatments. Fourth, for outpatient types of services, in-person doctor visits might be substituted with online consultations. While some of these direct effects on utilization patterns might persist (such as increased use of online consultations) others are likely to be transitory. In sum, these disruptions will occur in the pandemic year and to some degree in years post the pandemic.

In addition, there might be some indirect effects. For example, cancellation and postponement of regular treatments might lead to exacerbations of health problems, which – in the long run – can result in higher spending. At the same time, cancellation of regular treatments might lead to a decrease in iatrogenic diseases, which might result in lower spending, both in the short and in the long run.[1]

The direction and size of these direct and indirect effects are empirical questions. Given the typical delay in insurance-claim registration and the uncertainty about how the pandemic will develop, it might take years before we know the exact answer to these questions. It is this uncertainty that complicates the calibration of risk adjustment payment weights and the calculation of payments to insurers.

2.2. How the pandemic can complicate calibration of payment weights

Risk adjustment models for health plan payment typically assume a linear relationship between healthcare spending and a predefined set of risk adjuster variables:

$$y_{i,t} = \beta_0 + \beta_1 x_{i,1,t} + \beta_2 x_{i,2,t} + \cdots + \beta_p x_{i,p,t} + \varepsilon_{i,t} \qquad (1)$$

with $y_{i,t}$ the value of healthcare spending for individual *i* in year *t*, $x_{i,j,t}$ the value for *i* on risk adjuster *j* (typically a dummy variable) in year t, β_0 a constant term, β_j the average effect of risk adjuster *j* on y (holding other risk adjusters fixed) and $\varepsilon_{i,t}$ the error term in year *t*. In practice, the number of risk adjuster variables *p* used in risk adjustment models runs into the tens or hundreds. Typical information on which these variables are based includes age, gender, socioeconomic factors and diagnoses (either from year *t* in case of concurrent models or from a prior period in case of prospective models).

Assuming (1) reflects the 'actual' relationship between risk adjusters and spending in year *t*, the goal of the regulator is to estimate this relationship as accurate as possible. The best estimation of β_j can be obtained by a regression of spending on risk adjusters using ***realized*** information from year *t* itself:

$$y_{i,t} = b_{0,t} + b_{1,t} x_{i,1,t} + b_{2,t} x_{i,2,t} + \cdots + b_{p,t} x_{i,p,t} + e_{i,t} \qquad (2)$$

[1] For a helpful description of the potential long-term implications of COVID-19 see https://hcldr.wordpress.com/2020/04/07/the-pandemics-4th-wave/

with $x_{i,j,t}$ the value for *i* on risk adjuster *j* in year *t* and b_j the estimated payment weight for x_j in year *t*. This regression inherently provides payment weights that accurately reflect the (assumed linear) relationship between risk adjusters and spending in the payment year. As far as we know, this procedure for estimating payment weights is only applied in Germany [11]. In other countries, the relationship between risk adjusters and spending in payment year *t* is estimated on a calibration sample from a previous period. If we assume a delay *d* (measured in years) between the calibration sample and the payment year, the regression model can be written as:

$$y_{i,t-d} = b_{0,t-d} + b_{1,t-d}x_{i,1,t-d} + b_{2,t-d}x_{i,2,t-d} + \cdots + b_{p,t-d}, x_{i,p,t-d} + e_{i,t-d} \qquad (3)$$

In these settings it is important to make the calibration sample 'representative' for payment year *t*. Any discrepancy in the relationship between risk adjusters (*x*) and spending (*y*) between the calibration sample and the population/year of interest can result in payment weights (*b*) that do not accurately reflect the relationship between risk adjusters and spending in the payment year (β) as presented in (1). Typical modifications of calibration samples to avoid such discrepancies include corrections for cost inflation, benefits package and composition of the population. Until recently, in most settings, (modified) calibration samples have been considered reasonably representative. Due to COVID-19, however, this might no longer be the case, at least not for the years to come. The direction and size of discrepancies between the calibration sample and the payment year is uncertain (as long as accurate data on the impact of COVID on spending is lacking) and might differ from setting to setting.

2.3. How the pandemic can complicate calculation of payments to insurers

In the second step of risk adjustment, the estimated payment weights are combined with enrollee characteristics to generate a prediction of spending for health plan *k* that forms the basis for the payment to *k* in year *t*. Both here and in our quantitative illustration we assume $\hat{y}_{k,t}$ is calculated as:

$$\hat{y}_{k,t} = \sum_{i \in k}(b_{0,t-d} + b_{1,t-d}x_{i,1,t} + b_{2,t-d}x_{i,2,t} + \cdots + b_{p,t-d}x_{i,p,t}) \qquad (4)$$

, that is the sum of predicted spending over all individuals enrolled in plan *k*. It is straightforward to see how 'biased' payment weights *b* from (3) would result in a 'biased' value of predicted spending, an argument already made in the previous section. For example, if $b_{0,t-d}$ resulting from (3) is higher (lower) than b_0 in (1) then payments to insurers might be too high (low), ceteris paribus. Note, however, that this is under the assumption that predicted spending is indeed calculated as in (4). In practice, other modalities are present in which predicted spending does not 'simply' equal the predicted value from the regression model – as in (4) – but results from a multiplication of a 'relative risk score' (derived from the regression model) and some form of mean spending or premiums in year *t*. For example, the payment models used under the Affordable Care Act in the U.S. combine a 'relative risk score' with the mean insurance premium in year *t* in the market of interest. Depending on their exact specification, such payment modalities might implicitly correct for disruptions in overall spending in the market.

Apart from the issue of distorted weights, biased payments can also result from disruptions in patterns of risk adjuster flags (*x*) directly. In several payment systems, risk adjuster flags are based on information from a prior period. Such risk adjusters are commonly referred to as 'prospective' risk adjusters, as opposed to 'concurrent' risk adjusters that are based on information from the payment year itself. Prospective risk adjusters are used in Germany (where Hierarchical Morbidity Groups are based on diagnoses from the prior year) and Switzerland (where Pharmaceutical Cost Groups are based on drugs from the prior year). The Netherlands go even further by using information from up to *five* consecutive prior years as a basis for spending-based risk adjusters (see Section 4).[2] With prospective risk adjusters, a disruption in patterns of utilization can influence payments in future years, *even when the payment weights b are 'unbiased'*. The next section provides a hypothetical illustration of these dynamics.

Examples of concurrent risk adjusters can be found in the U.S. under Medicare Advantage and the Affordable Care Act. With concurrent risk adjusters, irregularities in flags (*x*) might be less of a problem than with prospective risk adjusters since flags in year *t* are directly related to spending in year *t*. For example, a reduction in hospital treatments in *t* is likely to result in both fewer diagnostic flags and lower spending. This implies that for insurers with fewer diagnostic flags, both the risk adjusted payment and spending are likely to reduce. With prospective risk adjustment on the other hand, fewer flags based on information from year *t*-1 results in a reduction of the payment for year *t*, but not (necessarily) in a reduction of spending in year *t*, creating a gap between payments and spending.

[2] An argument for using prospective instead of concurrent risk adjusters is that the latter might not only correct for predictable spending variation but also for *un*predictable spending variation [12].

2.4. Key observations

In sum, we can conclude that disruptions in healthcare utilization and spending due to the COVID-19 pandemic can distort health plan payments, either via the estimation model or via the payment model.

3. A hypothetical illustration

This section provides a hypothetical quantitative illustration of how the pandemic can distort estimation of payments weights (step 1) and calculation of payments to insurers (step 2). To do so, we make some assumptions on how COVID-19 interferes with the regular pattern of risk adjuster flags and how this affects the estimation model and the payment model. It should be emphasized that our simulation is purely hypothetical. In practice, outcomes can be (very) different depending on contextual factors (such as the scope and depth of coverage), the specific risk adjustment methodology in place and the actual impact of the pandemic on spending and risk adjuster flags. The goal of our illustration is to give an impression of how and under what circumstances COVID-19 can distort payments to insurers.

The data available for our simulation comes from a Swiss insurance company and contains information on individual-level spending and risk adjuster flags for about 1.3 million individuals. Spending is presented in Swiss francs (CHF)[3] and comprises all insurance claims for services covered by the Swiss basic health insurance in 2016. Risk adjuster flags include age, gender, prior hospitalization and pharmaceutical-based cost groups (PCGs). For the purpose of our illustration, we primarily focus on 'prior hospitalization'.[4] This risk adjuster takes the form of a dummy variable indicating whether a person spent at least three consecutive nights in a hospital or nursing home in 2015. Combined with spending in 2016, this risk adjuster allows us to run a series of simulations for a simple prospective risk adjustment model. In a later step, we extent this risk adjustment model with the risk adjuster 'yes/no PCG' to show how a distortion of one risk adjuster can be mitigated by the presence of another. PCGs in our dataset are derived from use of specific prescribed drugs in 2015. Although the PCG-classification distinguishes multiple diseases, we restrict our analysis to 'yes/no PCG', again for simplification.

[3] 1 CHF = 0.93 euro on November 5, 2020.

[4] As we will show later on, the main point of our illustration is that a disruption of flags regarding the risk adjuster 'prior hospitalization' (either in the calibration sample or in the payment year itself) can distort payments to insurers. For making this point it is not necessary to include other risk adjusters like age and gender.

Table 1 presents some key statistics that provide a starting point for our simulation. Mean per person spending in the population equals CHF 3,844. For the purpose of our analysis, per person spending is annualized by dividing observed spending by the fraction of the year an individual was enrolled, a common procedure in risk adjustment. All figures presented in this section are weighted by this fraction.

Table 1. Descriptive statistics

	Frequency in year t		**Mean spending year t (CHF)**	**% prior hosp.**		**% PCG**	
	Insured years	**% of total population**		**no**	**yes**	**no**	**Yes**
Descriptive statistics							
Total population	1,255,321	100	3,844				
Prior hospitalization = no	1,187,026	94.6	3,177			81.0	19.0
Prior hospitalization = yes	68,295	5.4	15,431			38.8	61.2
PCG = no	988,414	78.7	2,059	97.3	2.7		
PCG = yes	266,908	21.3	10,453	84.3	15.7		
Insurers with roughly identical portfolios							
Insurer 1	627,716	50.0	3,843	94.6	5.4	78.7	21.3
Insurer 2	627,605	50.0	3,845	94.6	5.4	78.7	21.3
Insurers with non-identical portfolios							
Insurer 1	627,766	50.0	4,239	91.3	8.7	77.4	22.6
Insurer 2	627,556	50.0	3,448	97.8	2.2	80.1	19.9

Note: 1 CHF = 0.93 euro (on November 5, 2020).

Table 1 also shows the mean per person spending in two mutually exclusive groups based on whether an individual is flagged by the risk adjuster 'prior hospitalization'. Relative to the total population, 5.4% is flagged by this risk adjuster with mean per person spending of CHF 15,431. For the complementary group (94.6%) mean per person spending equals CHF 3,177. The difference in mean spending between these groups indicates that prior hospitalization is indeed predictive of spending in year *t*. A similar observation can be made for the risk adjuster 'PCG = yes': for the group of enrollees flagged by this risk adjuster (21.3% of the population) mean spending equals CHF 10,453 per person per year; for the complementary group (78.7%) mean spending equals CHF 2,059 per person per year.

In terms of payment fit, we are primarily interested in the under/overpayment of insurers (see introduction). As will be described in more detail below, we simulate under/overpayment for two hypothetical insurers that each cover 50% of the population. Two market scenarios are taken into account. In the first scenario insurers have roughly identical portfolios. For this scenario, we randomly assign 50% of the population to insurer 1 and the other 50% to insurer 2. In the second scenario, insurer 1 has a disproportionate share of high-risk people while insurer 2 has a disproportionate share of low-risk people. For this scenario, we randomly assign 80% of the group with ‘prior hospitalization = yes’, together with a random 48.3% of the group with ‘prior hospitalization = no’ to insurer 1. All individuals not assigned to insurer 1 (i.e. 20% of the group with ‘prior hospitalization = yes’ and 51.7% of the group ‘prior hospitalization = no’) are assigned to insurer 2. Table 1 presents the portfolio size and mean spending for the two insurers per scenario. The third and fourth columns to the right show that with non-identical portfolios insurer 1 indeed has more enrollees with ‘prior hospitalization = yes’ than non-random insurer 2, resulting in a difference in mean per person spending of about 800 CHF. With non-identical portfolios insurer 1 also has a disproportionate share of people with ‘PCG = yes’.

3.1. Distortion of the calibration sample

As a first step in our illustration, we show how biased payment weights can affect outcomes for individual insurers. This illustration can be framed as follows. Assume that spending and risk adjuster flags in our dataset are perfectly representative for the payment year. In the calibration step, the objective of the regulator is to find the ‘right’ payment weights to be used in formula (4). To estimate these weights, the regulator relies on a calibration sample from a previous period. We simulate three scenarios regarding this calibration sample. For each scenario we estimate payment weights for a risk adjustment model with two dummy variables (and no intercept): one dummy for ‘prior hospitalization = yes’ and one for ‘prior hospitalization = no’. The model takes the form of a least squares regression of (annualized) spending on these two dummy’s (with a weight for duration of enrollment).

The first scenario (W0, with ‘W’ for weights) assumes that the calibration sample is *not* distorted by the pandemic. More specifically, we assume there is no discrepancy between the calibration sample and the payment year. As shown in Table 2, the payment weights in this scenario simply equal the mean per person spending for the respective groups that were already presented in Table 1. As indicated by the R-squared, the risk adjustment model explains 6.9% of the variance in spending in our dataset.

The second scenario (W1) assumes the calibration sample is distorted by the pandemic in the following way: flags for ‘prior hospitalization = yes’ decreased by a random 10%. For this scenario we randomly assign 10% of the group with ‘prior hospitalization = yes’ to the group with ‘prior hospitalization = no’, which is reflected in the frequency of these flags. Because of this distortion, the group with no prior hospitalization becomes somewhat more heterogeneous, resulting in an increase of the payment weight for this group and a slight reduction in the R-squared of the estimation model (see Table 2).

Table 2. Payment weights in three scenarios regarding the calibration sample

	Frequency of flags in calibration sample (%)	**Payment weight (CHF)**
Scenario W0: No distortion of calibration sample		
≥3 nights hospitalized in t-1 = no	94.6	3,177
≥3 nights hospitalized in t-1 = yes	5.4	15,431
R-squared	0.069	
Scenario W1: random distortion of calibration sample		
≥3 nights hospitalized in t-1 = no	95.1	3,246
≥3 nights hospitalized in t-1 = yes	4.9	15,456
R-squared	0.068	
Scenario W2: non-random distortion of calibration sample		
≥3 nights hospitalized in t-1 = no	95.1	3,192
≥3 nights hospitalized in t-1 = yes	4.9	16,559
R-squared	0.068	

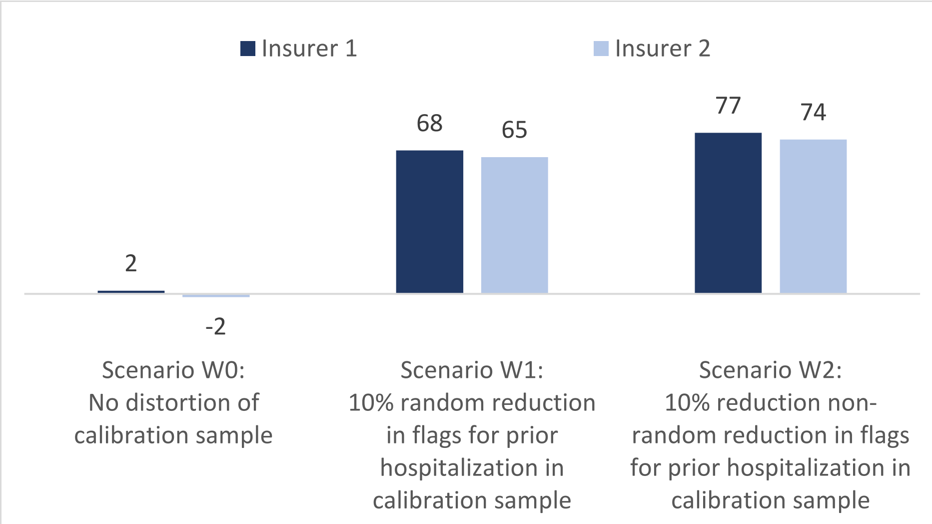
The third scenario (W2) contains another type of distortion: like in W1, flags for prior hospitalization in the calibration sample decreased by 10%, but this time the drop is non-random. More specifically, the decrease in flags took place among the relatively low-risk people within the group ‘prior hospitalization = yes’. In order to make a distinction between the relatively low risks and relatively high risks in the group ‘prior hospitalization = yes’ we calculated median spending in this group in 2015. Those with below-median spending were labelled as relatively low-risk people. For a random 20% of these low-risk people we changed the flag ‘prior hospitalization = yes’ to ‘prior hospitalization = no’. As shown in Table

2, this results in a decrease in flags for prior hospitalization from 5.4% to 4.9%, similar to scenario W1. Contrary to W1, however, this drop makes the group with ‘prior hospitalization = yes’ more homogenous, which is reflected in a higher weight for this group compared to W0 and W1. In scenario W2, the group ‘prior hospitalization = no’ is more heterogeneous than in W0, but less heterogeneous than in W1, resulting in a payment weight in between the corresponding weights in W0 and W1.

The main takeaway from Table 2 is that estimated payment weights are driven by the composition of the calibration sample (in terms of spending and risk adjuster flags). Our primary question for what follows is how the differences in payment weights in Table 2 affect outcomes for individual insurers. To simulate these outcomes we plug in the estimated payment weights in formula (4) while keeping ‘risk adjuster flags’ constant (that is, we use the flags as present in our original dataset).

Figure 1 shows the mean financial result under scenarios W0-W2 for two insurers with roughly identical portfolios. The mean financial result is calculated as the mean per person risk adjusted payment minus the mean per person actual spending in our dataset. Three important observations can be made. First, without a distortion (scenario W0, in which the calibration sample is perfectly representative for the payment year), the financial result for both insurers equals nearly zero. Second, with a random distortion (scenario W1, in which the calibration sample includes a random 10% decrease in flags for ‘prior hospitalization = yes’), both insurers are confronted with a positive financial result. The explanation is that – due to the distortion in the calibration sample – the payment weight for the flag ‘prior hospitalization = no’ is too high. Third, with a non-random distortion (scenario W2, in which the calibration sample includes a non-random decrease in flags for ‘prior hospitalization = yes’), the overpayment for both insurers is even larger than in W1. The reason is that in W2, both the payment weight for ‘prior hospitalization = yes’ and ‘prior hospitalization = no’ are too high. Despite the average overpayment in the market, however, the distortions in the calibration sample do not alter the level playing field: in all three scenarios the overpayment for the two insurers is more or less similar.

Figure 1. Mean financial result in CHF per person per year for insurers with ***roughly identical*** portfolios



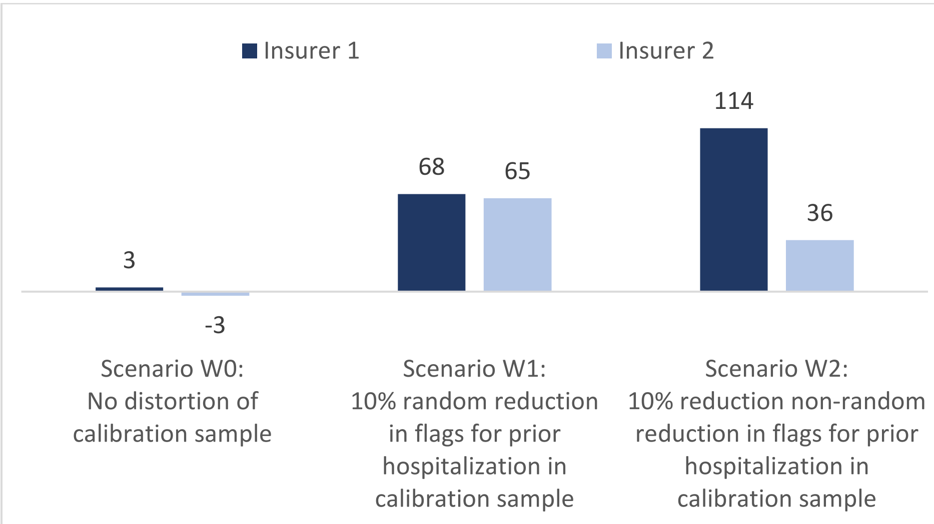
Note: Both insurers cover a random 50% percent of the population. See 'roughly identical insurers' in Table 1.

Like Figure 1, Figure 2 presents the mean financial result under scenarios W0-W2. This time, however, the insurers are assumed to have non-identical portfolios. More specifically, insurer 1 covers more people with prior hospitalization than insurer 2. Again three observations can be made. First, without a distortion (scenario W0), the mean financial result is nearly zero for both insurers, despite the substantial differences in mean spending between the two portfolios (see rows for 'non-random insurers' in Table 1). The explanation is that the risk adjustment model perfectly compensates insurer 1 (insurer 2) for the overrepresentation (underrepresentation) of people with prior hospitalization. Second, with a random distortion (scenario W1, in which the calibration sample includes a random 10% decrease in flags for 'hospitalization=yes'), both insurers are confronted with a positive financial result. Again, the explanation is that – due to the distortion in the calibration sample – the payment weight for the flag 'prior hospitalization = no' is too high. Both insurers benefit from this since both cover many people with 'prior hospitalization = no' (prevalence of this flag is 91% for insurer 1 and 98% for insurer 2).[5] The most remarkable outcome in Figure 2 is for scenario W2. In this scenario (where the calibration sample includes a non-random decrease in flags for 'prior hospitalization = yes'), the overpayment for insurer 1 is substantially higher than for insurer 2. The reason is to be found in the higher prevalence of

[5] Surprisingly, the effect is more or less the same for insurer 1 and 2, despite the 7 percent point difference (98%-91%) in prevalence of 'prior hospitalization = no'. The explanation for this counterintuitive effect is that the payment weight for 'prior hospitalization = yes' also increases somewhat compared to scenario W0 (see Table 2; caused by random variation) which mitigates the net effect. A second explanation is that in relative terms the difference in prevalence of 'prior hospitalization =no' between the two insurers is rather small.

‘prior hospitalization = yes’ for insurer 1 (8.7%) compared to insurer 2 (2.2%). Due to this difference, insurer 1 benefits more from the ‘biased’ payment weight for this risk adjuster than insurer 2.[6]

Figure 2. Mean financial result in CHF per person per year for insurers with ***non-identical*** portfolios



Note: Both insurers cover 50% percent of the population. Insurer 1 covers more people with prior hospitalization than insurer 2. See rows for ‘non-identical insurers’ in Table 1.

In sum, the illustration above leads us to two important conclusions. First, biased payment weights resulting from distortions in the calibration sample can result in overpayment or underpayment of the entire market.[7] (Although our scenarios W1 and W2 resulted in overpayment, distortions can go in the other direction as well. This might be the case, for instance, when disease flags increase instead of decrease, e.g. due to an increase of COVID-19 related treatments.) Second, to the extent that insurers’ portfolios differ in terms of composition, biased payment weights can distort the level playing field.

3.2. Distortion of flags used in the payment model

As a second step in our quantitative illustration, we show how a distortion of ‘risk adjuster flags’ can affect outcomes for individual insurers through the payment model. This illustration can be framed as follows. Assume that payment weights are perfectly representative for the relationship between risk adjusters and spending in the payment year. In other words, we assume that formula (4) includes the

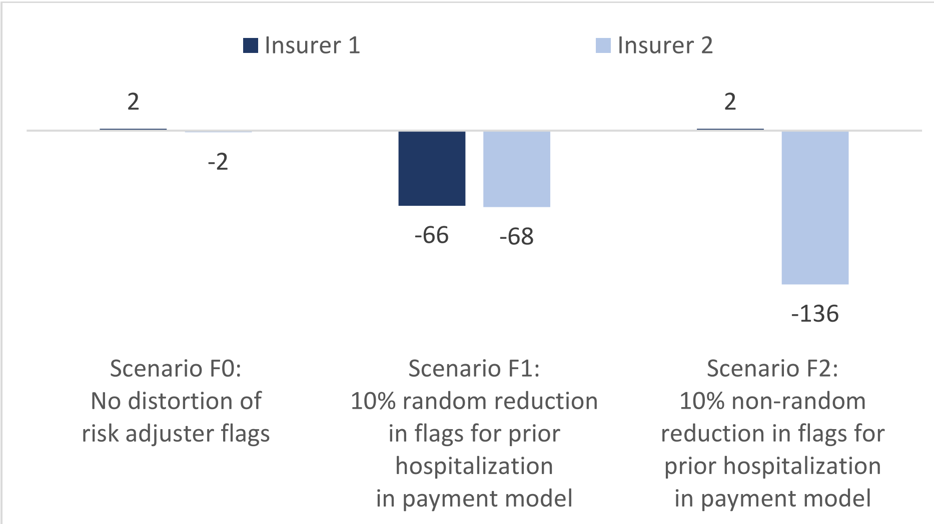
[6] Note that both insurers also benefit from the ‘biased’ payment weight for ‘prior hospitalization =no’.
[7] Note that our simulation assumes an ex-ante fixed budget. Payment models that take into account actual premiums or spending might to some extent correct for over/underpayment of the entire market.

‘right’ set of payment weights b. Our focus here is on the set of risk adjuster flags (x) for which we simulate three scenarios. These scenarios differ in terms of the risk adjuster flags (x) used for calculating payments while keeping payment weights (b) equal. The first scenario (F0, with ‘F’ for flags) assumes that the risk adjuster flags used for calculation of plan payments are *not* distorted by the pandemic. In this scenario, we use the flags as present in our dataset. The second scenario (F1) assumes a random distortion of flags used for the calculation of plan payments: flags for prior hospitalization decreased by a random 10%. For this scenario, we randomly assign 10% of the group with ‘prior hospitalization = yes’ to the group with ‘prior hospitalization = no’. The third scenario (F2) assumes a non-random distortion of flags used for calculation of payments: like in scenario F1, flags for prior hospitalization decreased by 10%, but this time the drop is non-random. More specifically, the decrease took place in the portfolio of *one* insurer (for which the number of flags for ‘prior hospitalization = yes’ decreased by 20%). In all three scenarios, the payment weight for ‘prior hospitalization = no’ equals 3,177 CHF and the weight for ‘prior hospitalization = yes’ equals 15,431 CHF (i.e. the undistorted weights under scenario W0 in Table 2).

In what follows, we are interested in how disruptions in risk adjuster flags distort payments to insurers. Similar to the procedure used in the previous section we calculate under/overpayments for insurers as the mean per person risk adjusted payment minus the mean per person spending. Although (in F1 and F2) risk adjuster flags for yes/no hospitalization in year *t*-1 are disrupted, we assume that spending in year *t* is not. Again, we run our simulations for random portfolios and non-random portfolios.

Figure 3 shows the outcomes of our simulation for two random portfolios. As expected, the mean financial result for insurers in scenario F0 (no distortion in risk adjuster flags) equals nearly zero. Slight deviations from zero are due to random variation. A decrease in flags for ‘prior hospitalization = yes’ (scenarios F1 and F2) results in lower payments to plans. When levels of actual spending remain the same (as we assume here) both insurers are confronted with an underpayment. Not surprisingly, a decrease in flags for ‘prior hospitalization = yes’ hits some insurers harder than others as this decrease is concentrated in specific plans (scenario F2). To some extent, scenario F2 might occur when insurers have relatively high market shares in regions that have seen relatively high COVID-19 incidence rates.

Figure 3. Mean financial result in CHF per person per year for insurers with ***roughly identical*** portfolios

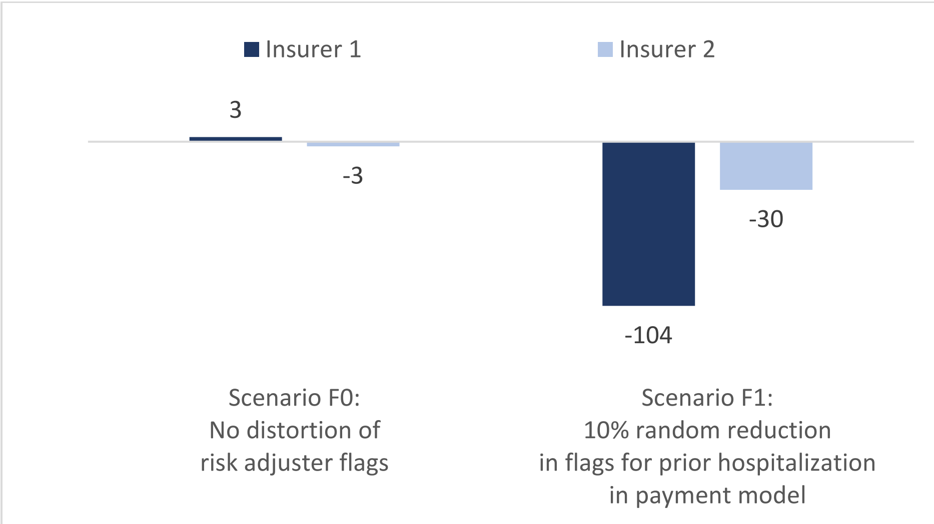


Note: Both insurers cover a random 50% percent of the population. See 'roughly identical insurers' in Table 1.

Figure 4 shows the outcomes of our simulation for two ***non-identical*** portfolios. Again, as expected, the mean financial result for insurers in scenario F0 (no distortion in risk adjuster flags in the payment model) equals nearly zero, implying that the risk adjustment model fully compensates for the difference in mean spending in the two plans. Similar to Figure 3, a decrease in flags for 'prior hospitalization = yes' (scenarios F1) results in lower payments to plans. This time, however, the effects for insurers are different. Although, the mean underpayment in the market is similar to that in Figure 3, insurer 1 is now losing more money than insurer 2. The explanation is that insurer 1 has more enrollees flagged by the risk adjuster 'prior hospitalization =yes' (8.7% of his portfolio) than insurer 2 (2.2%). Consequently, a 10% reduction in flags in the payment model hurts insurer 1 more than insurer 2.

For simplicity, Figure 4 does not include F2. It is easy to see how a non-random decrease in flags for 'prior hospitalization = yes' would change the results compared to scenario F1: if flags would decrease for one insurer but not for the other, the underpayment would be concentrated with that insurer (see Figure 3), irrespective of whether insurers have identical portfolios or not.

Figure 4. Mean financial result in CHF per person per year for two insurers with ***non-identical*** portfolios



Note: Both insurers cover 50% percent of the population. Insurer 1 covers more people with prior hospitalization than insurer 2. See rows for 'non-identical insurers' in Table 1.

In sum, Figures 3 and 4 lead us to two important conclusions. First, distortions in risk adjuster flags in the payment model can result in underpayment or overpayment of the entire market.[8] (Although our scenarios F1 and F2 resulted in underpayment, distortions can go in the other direction as well, e.g. when disease flags increase instead of decrease as a result of more COVID-19 related treatments.) Second, to the extent that distortions in risk adjuster flags are more severe for some insurers than for others (e.g. due to non-random portfolios and/or relatively high COVID-19 incidence rates in regions with concentration of specific insurers), disruptions in risk adjuster flags can alter the level playing field.

3.3. How other risk adjusters might rescue the risk adjustment system

So far, our simulations assumed that the risk adjustment system includes just one risk adjuster, i.e. 'yes/no prior hospitalization'. In practice, risk adjustment systems are much more sophisticated and often include a wide range of risk adjusters. An interesting question is how the outcomes of our simulations would change if we included additional risk adjusters. To answer this question we repeated our simulation for a risk adjustment system that includes both 'yes/no prior hospitalization' and 'yes/no PCG'. Table 3 summarizes the outcomes of this simulation and compares these outcomes with those of

[8] Note that our simulation assumes an ex-ante fixed budget. Payment models that take into account actual premiums or spending might to some extent correct for over/underpayment of the entire market.

our previous simulation. More specifically, Table 3 shows the overall mean per person financial result (which indicates the mean per person under/overpayment across the entire market) and the absolute difference in mean per person financial result between insurer 1 and insurer 2 (which indicates the extent to which the level playing field is being altered). The results in the columns 'Without PCG' replicate the outcomes of our previous simulation with only one risk adjuster. The columns 'With PCG' show the outcomes for the risk adjustment system with two risk adjusters; in all scenarios we assume that flags for the risk adjuster 'yes/no PCG' are <u>not</u> distorted by the COVID-19 pandemic.

Table 3. Outcomes per scenario without/with risk adjuster 'PCG' added to the risk adjustment system

Scenario	**Overall mean per person financial result across the market (CHF)**		**Absolute difference in mean per person financial result between insurer 1 and 2 (CHF)**	
	Without PCG	**With PCG**	**Without PCG**	**With PCG**
Roughly identical insurers:				
W0	0	0	3	4
W1	67	50	3	4
W2	75	58	3	4
Non-identical insurers				
W0	0	0	6	7
W1	67	50	3	6
W2	75	58	78	72
Roughly identical insurers:				
F0	0	0	3	4
F1	-67	-50	1	2
F2	-67	-50	137	105
Non-identical insurers:				
F0	0	0	6	7
F1	-67	-50	73	53

Note: Results in columns 'Without PCG' can be derived from Figures 1-4. For example, Figure 1 indicates that in scenario W0 the mean per person financial result (overall) equals zero and that the absolute difference in mean per person financial result between the two insurers equals 4 euro. Small differences are due to rounding.

Figures 1-4 have shown that scenarios W1, W2, F1 and F2 result in under- or overpayment across the market (see also column 2 of Table 3). Column 3 of Table 3 shows that these under/overpayments shrink when the risk adjuster ‘PCG’ is added. A similar observation can be made for the absolute difference in mean per person financial result between the insurers: in the scenarios ‘Non-identical insurers: W2’, ‘Roughly identical insurers: F2’ and ‘Non-identical insurers: F1’ this difference is substantially smaller in the risk adjustment system ‘With PCG’ (column 5) than in the system ‘Without PCG’ (column 4). In other words, the distortions of both the estimation model and the payment model are smaller for a system with both ‘yes/no prior hospitalization’ and ‘yes/no PCG’ than for a system with only the former. The explanation can be found in Table 1 (in the four columns to the right): people with prior hospitalization have a relatively high probability of being flagged by a PCG, and vice versa. The correlation between the two risk adjusters in our dataset turns out to be .237. Due to this correlation, adding ‘yes/no PCG’ as a risk adjuster reduces the payment weight for the risk adjuster ‘yes/no prior hospitalization’ implying that a distortion of this risk adjuster results in less damage. This brings us to an interesting observation: distortions from COVID-19 via a specific risk adjuster, say x_1 (e.g. ‘yes/no prior hospitalization’), are smaller as the risk adjustment system includes additional risk adjusters that are correlated with x_1 AND not (substantially) affected by the pandemic. The extent to which ‘other risk adjusters can rescue a risk adjustment system’ is an empirical question depending on the features of that system and the correlation between these other risk adjusters and those distorted by COVID-19.

4. An exemplary case: Health plan payment in the Netherlands

In this section we qualitatively discuss the impact of the COVID-19 pandemic for a real-world risk adjustment system, i.e. the one used for health plan payment in the Netherlands. We believe this system forms an interesting case as it shares many commonalities with systems used in other countries and maybe useful for exposing the vulnerabilities of real-world payment systems to disruptions caused by the pandemic. Below, we will first describe some essential features of the Dutch system.

4.1. Risk adjustment in the Dutch basic health insurance

The Dutch risk adjustment system includes a diverse set of demographic and socioeconomic variables including age, gender, regional characteristics, source of income, education, household size and income. In addition, the system also includes a wide range of more direct indicators of health based on utilization and spending from prior years; these include Diagnoses-based Cost Groups (DCGs), Pharmacy-based Cost Groups (PCGs), Physiotherapy-Diagnoses Cost Groups (PDCGs), Durable Medical Equipment Cost

Groups (DMECGs) and Multiple-year High Cost Groups (MHCGs). For a detailed description of these risk adjusters we refer to Van Kleef et al. (2018). For our analysis below it is important to note that most of the demographic and socioeconomic variables are used concurrently while the health indicators are used prospectively. More specifically, DCGs, PCGs, PDCGs and DMECGs are based on information from the prior year; MHCGs are based on information from up to five consecutive prior years.

To estimate the payment weights for year *t*, the regulator relies on a calibration sample with spending from year *t*-3. This calibration sample includes all individuals with a basic health insurance in *t*-3. Before the estimation of payment weights, the calibration sample is modified to make it representative for the year of interest (year *t*). Typical modifications include corrections for cost inflation and changes in the benefits package between year *t*-3 and year *t* and sample-rebalancing for changes in risk adjuster flags between *t*-3 and *t*. After these modifications (and thus taking into account the projected spending levels for year *t*), payment weights are estimated by a least-squares regression of spending on risk adjusters.

Calculation of payments to plans consists of a series of steps in which insurers first receive a provisional payment based on their ***projected*** portfolio in year *t*. In next steps, this provisional payment is adjusted for ***actual*** portfolio in year *t* as this becomes known. For simplification, we will primarily focus on the final payment to insurers (and not the provisional payments, which are just hypothetical). For the calculation of final payments the regulator combines the estimated payment weights with the actual risk adjuster flags from year *t* to year *t*-5. For example, age and gender are based on year *t*, while PCGs and DCGs are based on information from year *t*-1. For details we refer to Van Kleef et al. (2018). In practice, some ex-post corrections are applied to plan payments. Although we do not take into account these corrections here, some relate to the policy measures discussed in the next section.

Given these features of the Dutch risk adjustment system, the COVID-19 pandemic is expected to complicate both the calibration of payment weights (Section 4.3) and the calculation of final payments to plans (Section 4.4). Before we discuss these complications, however, we first give an overview of the impact of COVID-19 on patterns of spending and risk adjuster flags in the Dutch context (Section 4.2).

4.2. Disruptions of spending and risk adjuster flags

Table 4 indicates for which years disruptions in spending might complicate the Dutch risk adjustment system. The first column shows the year for which payments to plans are calculated. The second column presents the year of spending used for calibration. The last two columns indicate whether the COVID-19 pandemic is expected to distort spending patterns, either via the payment model or the estimation model. Two scenarios are taken into account, one in which COVID-19 distorts spending patterns in 2020 'only' and one in which COVID-19 distorts spending in both 2020 and 2021 (but not in later years). One observation is that discontinuities in spending will 'hit' the risk adjustment system twice; first through the payment model and then – three years later – through the estimation model. Another observation is that complications get bigger when spending continues to be distorted in 2021. In this case, disruptions in spending will also 'hit' the risk adjustment system in 2021 and 2024. This scenario is not unlikely given the possibility of new waves in COVID-19 infections or the possibility that in 2021 doctors/patients catch up some of the treatments that were postponed in 2020. Note that distortions in spending patterns might also affect the risk adjuster 'Multiple-year high spending', an affect that we will discuss below together with the effects of COVID-19 on other risk adjuster variables.

Table 4. Years in which disruptions in spending complicate the Dutch risk adjustment system

Year for which payments are calculated (year t)	**Year of spending used in estimation model (year t-3)**	**Yes/no impact on risk adjustment given that the COVID-19 pandemic disrupts spending in…**	
		2020 only	**2020-2021**
2020	2017	Yes	Yes
2021	2018	No	Yes
2022	2019	No	No
2023	2020	Yes	Yes
2024	2021	No	Yes
2025	2022	No	No

Table 5 indicates for which years distortions of risk adjuster flags might complicate the Dutch risk adjustment system. Like in Table 4, the first column shows the year for which payments to plans are calculated. The second column presents the years on which risk adjuster flags are based in the ***payment model*** and the thirds columns shows these years for the ***estimation model***. The last two columns indicate whether COVID-19 is expected to distort patterns in risk adjuster flags, either in the payment model or the estimation model. Again, two scenarios are taken into account, one in which COVID-19 disrupts spending patterns in 2020 (but not in 2021 and later years) and one in which COVID-19 disrupts spending in both 2020 and 2021 (but not in later years). A first observation is that – even when the COVID-19 pandemic 'only' disrupts utilization in 2020 – the Dutch risk adjustment system will be confronted with complications for many years. The reason is two-fold. First, a disruption of utilization in a certain year hits the system twice due to the three-year time lag in data used for the estimation model. Second, a distortion in flags for spending-based risk adjusters in data from a certain year (e.g. 2020) will be carried along for multiple years due to the time span on which these risk adjusters are based. The nature and size of distortions in risk adjuster flags, however, will differ from year to year. Potential distortions include lower prevalence of elderly people (due to higher mortality rates among these people in 2020), fewer flags for specific morbidity indicators based on information from the year 2020 (due to a reduction in regular treatments in 2020), more flags for other specific indicators based on information from the year 2020 (due to an increase in COVID-19-related treatments), changes in composition of groups with multiple-year high spending (due to a distortion of spending patterns in 2020), changes in the composition of socioeconomic classes due to economic downturn. The direction and size of such discontinuities remain to be an empirical question and will heavily depend on the development of the COVID-19 pandemic and its effects over the next few years.

Table 5. Years in which distortions of risk adjuster flags complicate the Dutch risk adjustment system

Year for which payments are calculated	Years of risk adjusters used for payment model	Years of risk adjusters used for estimation model	Yes/no impact on risk adjustment given that the COVID-19 pandemic disrupts utilization/spending in…	
			2020 only	**2020-2021**
2020	2015-2020	2012-2017	Yes	Yes
2021	2016-2021	2013-2018	Yes	Yes
2022	2017-2022	2014-2019	Yes	Yes
2023	2018-2023	2015-2020	Yes	Yes
2024	2019-2024	2016-2021	Yes	Yes
2025	2020-2025	2017-2022	Yes	Yes
2026	2021-2026	2018-2023	Yes	Yes
2027	2022-2027	2019-2024	Yes	Yes
2028	2023-2028	2020-2025	Yes	Yes
2029	2024-2029	2021-2026	No	Yes
2030	2025-2030	2022-2027	No	No

4.3. Complications for the estimation model

Both the discontinuities in spending and risk adjuster flags will result in discrepancies between the calibration sample and the year for which payments are calculated. If not corrected for, these discrepancies might lead to biased payment weights which can result in under/overpayment of the market and/or distort the level playing field for insurers (see Sections 2 and 3 of this paper). For some years, however, the potential bias might be bigger than for other years. For example, the bias might be more severe for years in which DCGs, PCGs and other morbidity indicators are based on data from 2020 than for years in which 'just' the MYHC indicator is (partly) based on data from 2020. The challenge for the Dutch regulator will be to correct for potential discrepancies as accurately as possible. For the near future, however, this will be hard due to uncertainty about the effect of COVID-19 on patterns in spending and risk adjuster flags and uncertainty about the development of the pandemic. Therefore, additional measures might be needed to mitigate potential under/overpayments, both at the market level and the level of insurers. In the next section, we will discuss a series of possible measures.

4.4. Complications for the payment model

As we have seen in Sections 2 and 3, discontinuities in spending and risk adjuster flags can also affect the payment model, keeping payment weights constant. For the payment model in 2020, for instance, risk adjuster flags are based on the year 2019. Although spending on regular care in 2020 is likely to be lower than anticipated, payments will more or less remain at the anticipated levels, resulting in an overpayment across the market. For the payment model of 2021, a different situation might occur: while spending on regular care might increase to pre-pandemic levels or beyond, frequencies of DCG flags – which are based on the year 2020 – might be lower due to a reduction of hospital care in 2020, resulting in an underpayment of the market and a distortion of the level playing field given that DCG flags are non-randomly distributed across plans (a situation comparable to Figure 4 of this paper). Similar to discontinuities in spending, the direction and size of discontinuities in flags are uncertain, at least for the near future. Therefore, over the next few years additional measures might be needed.

5. Discussion

Our analysis has shown how the COVID-19 pandemic can distort risk adjustment systems. Through disruptions in spending and risk adjuster flags the pandemic can 'hit' both the estimation and the payment model. Our quantitative example has shown how and under what circumstances these disruptions can lead to under/overpayment of the market and distort the level playing field. Our discussion of the potential impact COVID-19 on the Dutch risk adjustment system has shown that the distortions can continue for many years, even when the pandemic would be over in 2021.

We believe that our main findings are not only relevant for the Netherlands, but also for other countries. One of the main takeaways from this paper is that any system relying on a calibration sample from a previous period will be confronted with a potential distortion of estimated payment weights due to discrepancies between the calibration sample and the payment year. For example, this is true for the risk adjustment systems in Switzerland and specific sectors in the United States such as Medicare Advantage and the Marketplaces under the Affordable Care Act. Another main takeaway is that in any system with prospective risk adjusters, discontinuities in flags might complicate the payment model since these discontinuities (e.g. fewer diagnostic flags based on data from year t-1) do not necessarily follow spending patterns (in year t). Prospective risk adjusters are used in Germany and Switzerland. A third finding is that distortions from COVID-19 via a specific risk adjuster, say x_1 (e.g. 'yes/no prior hospitalization'), are smaller as the risk adjustment system includes additional risk adjuster variables that are correlated with x_1 *and* that are not affected by the COVID-19 pandemic. In this respect, distortions might be bigger in systems that heavily rely on hospital information as a basis for morbidity-adjusters than systems that use information from other healthcare settings as well.

Regulators could consider two general options to mitigate (the effects of) the impact of COVID-19 on risk adjustment: 1) redesign of risk adjustment methodology and/or 2) supplementary risk sharing measures. An example of the first option is to calibrate payment weights in retrospect by using realized spending and risk adjuster flags. This methodology, which is the standard procedure in Germany, makes sure that payment weights accurately reflect actual relationships between spending and risk adjusters in the year of interest (and thus avoids issues related to discrepancies between the calibration sample and year of interest). Instead of a full retrospective estimation model, regulators could consider modifying payment weights only for risk adjusters that are potentially problematic. Another example is to switch from prospective to concurrent risk adjusters. Concurrent models, as applied in risk adjustment systems under the Affordable Care Act and in Medicare Advantage, might mitigate under/overpayment due to distortions in risk adjuster flags since these distortions are likely to go hand in hand with distortions in spending. In a concurrent system, a reduction in diagnoses, for instance, is likely to result in both fewer diagnostic flags and (thus lower values of predicted spending) and lower values of actual spending. In a prospective system changes in payments and spending are less likely to go hand in hand. A third example of risk adjustment redesign is to simply 'keep out' problematic years, e.g. by not using calibration samples with crucial information (i.e. spending and/or risk adjuster flags) from 2020 and excluding the year 2020 from the construction of risk adjusters such as 'diagnoses cost groups' and 'multiple-year high spending' as used in the Dutch risk adjustment model.

An example of the second option is risk sharing between insurers and the regulator [13-14]. A possible measure could be, for instance, that the regulator subsidizes the market in case of an overall underpayment and to 'take back' money in case of an overall overpayment. Subsidies and repayments could be calculated in proportion to the initial risk adjusted payment. Such a form of risk sharing will be implemented in the Netherlands for 2021 to protect both the regulator and the market against a mismatch between payments and spending. Another modality is to apply a risk corridor for individual insurers in a way that they are subsidized for losses greater than X and/or cut for profits greater than Y.

This modality can be effective in protecting individual insurers from excessive financial risk and might mitigate distortions of the level playing field. Risk corridors can be applied either between insurers and the regulator (which implicitly protects the regulator and the market for excessive under/overpayment) or among insurers (leaving market-level under/overpayments intact). A third risk sharing option is to 'carve out' specific types of spending that are likely to be distorted. Other forms of risk sharing such as outlier-risk sharing might be less effective tools to deal with the impact of COVID-19.

Another potentially interesting form of risk sharing is between insurers and providers. To the extent that the pandemic reduces spending on regular types of care, surpluses on the side of insurers might help to compensate losses on the side of providers and vice versa. Such risk sharing can be organized by the regulator or the market itself. In The Netherlands, for instance, insurers and providers themselves agreed to move funds from insurers to providers to safeguard availability of healthcare services. The occurrence of COVID-19 might lead societies to reconsider the risks associated with a pandemic: should such risks be borne by insurers and/or healthcare providers or should it be borne by the regulator?
In sum, we conclude that discontinuities in spending and risk adjuster flags due to the COVID-19 pandemic can distort risk adjustment of health plan payments. Our analysis has shown how the direction and size of potential distortions depend on 1) the impact of COVID-19 on patterns in utilization and spending, 2) specific features of the risk adjustment methodology in place, 3) the distribution of risk types across plans, and 4) the extent to which insurers are unequally affected by the pandemic. As more data on the effects of COVID-19 on utilization and spending becomes available, regulators will be better equipped to deal with potential distortions. Given the commonalities in risk adjustment systems worldwide, we believe there is a lot to learn from international experiences over the next years.

References

1. Ellis, R.P., Rose, S., Martin, B.: 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. Elsevier. (2018)

2. Van Kleef, R.C, Eijkenaar, F., Van Vliet, R.C.J.A., Van de Ven W.P.M.M.: Health Plan Payment in the Netherlands. In: McGuire T.G., Van Kleef R.C. (eds.) Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice, pp. 397-429. Elsevier. (2018)

3. Schmid, C. P., Beck, K., Kauer, L.: Health Plan Payment in Switzerland. In: McGuire T.G., Van Kleef R.C. (eds.) Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice, pp. 453-489. Elsevier. (2018)

4. Layton, T.J., Montz, E., Shepard, M.: Health Plan Payment in U.S. Marketplaces: Regulated Competition with a Weak Mandate. In: McGuire T.G., Van Kleef R.C. (eds.) Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice, pp. 491-522. Elsevier. (2018)

5. McGuire, T.G., Newhouse J.P.: Medicare Advantage: Regulated Competition in the Shadow of a Public Option. In: McGuire T.G., Van Kleef R.C. (eds.) Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice, pp. 563-594. Elsevier. (2018)

6. Beck, K., Trottmann, M., Zweifel, P.: Risk Adjustment in Health Insurance and its Long-term Effectiveness. Journal of Health Economics 29: 489-98 (2010)

7. Buchner, F., Göpffarth, D., Wasem, J.: The new risk adjustment formula in Germany: Implementation and first experiences. Health Policy 109: 253-62. (2013)

8. Van Veen, S.H.C.M., Van Kleef, R.C., Van De Ven, W.P.M.M., Van Vliet, R.C.J.A.: Improving The Prediction Model Used in Risk Equalization: Cost and Diagnostic Information from Multiple Prior Years. European Journal of Health Economics 16: 201-218. (2015)

9. Kautter, J., Pope, G.C., Ingber, M., Freeman, S., Patterson, L., Cohen, M., Keenan, P.: The HHS-HCC Risk Adjustment Model for Individual and Small Group Markets Under the Affordable Care Act. Medicare & Medicaid Research Review 4: E1-E11. (2014)

10. Van de Ven, W.P.M.M., Van Kleef, R.C., Van Vliet, R.C.J.A.: Risk Selection Threatens Quality Of Care For Certain Patients; Lessons From Europe’s Health Insurance Exchanges. Health Affairs 34: 1713-1720. (2015)

11. Wasem, J., Buchner, F., Lux, G., Schillo, S. (2018). “Health Plan Payment in Germany,” In: McGuire T.G., Van Kleef R.C. (eds.) Risk Adjustment, Risk Sharing and Premium Regulation in Health Insurance Markets: Theory and Practice, pp. 295-329. Elsevier. (2018)

12. Van de Ven, W.P.M.M., Ellis, R.P.: Risk adjustment in competitive health insurance markets.” In: Culyer A.J., Newhouse, J.P. (eds.) Handbook of Health Economics, pp. 755-845. Elsevier. (2000)

13. Layton, T., McGuire, T.G., Sinaiko, A.S.: Risk Corridors and Reinsurance in Health Insurance Marketplaces: Insurance for Insurers. American Journal of Health Economics 2: 66-95. (2016)

14. Van Barneveld, E.M., Lamers, L.M., Van Vliet, R.C.J.A., Van de Ven, W.P.M.M.: Risk Sharing as a Supplement to Imperfect Capitation: A Tradeoff Between Selection and Efficiency. Journal of Health Economics 20: 147-168. (2001)

Erasmus University Rotterdam
Erasmus School of Health Policy & Management
Bayle gebouw
Burgemeester Oudlaan 50
3062 PA Rotterdam
T 010 408 8555
E communicatie@eshpm.eur.nl
W www.eur.nl/eshpm