Risk equalization in The Netherlands: an empirical evaluation


Richard C Van Kleef*, René CJA Van Vliet and Wynand PMM Van de Ven
Institute of Health Policy and Management, Erasmus University Rotterdam, The Netherlands
*Author for correspondence: Tel.: +31 10 4088 950 vankleef@bmg.eur.nl

The Netherlands relies on risk equalization to compensate competing health insurers for predictable variation in individual medical expenses. Without accurate risk equalization insurers are confronted with incentives for risk selection. The goal of this study is to evaluate the improvement in predictive accuracy of the Dutch risk equalization model since its introduction in 1993. Based on individual-level claims data (n = 15.6 million), we estimate the risk equalization models that have been successively applied in The Netherlands since 1993. Using individual-level survey data (n = 8735), we examine the average under-/overcompensation by these models for several relevant subgroups in the population. We find that in the course of years, the risk equalization model has been substantially improved. Even the current model (2012), however, does not eliminate incentives for risk selection completely. To achieve the public objectives, further improvement of the Dutch risk equalization model is crucial.

KEYWORDS: claims data • health insurance • risk equalization • risk selection • survey data

The Dutch basic health insurance includes a risk equalization model to compensate competing insurers for predictable variation in individual medical expenses between low-risk individuals (e.g., the healthy) and high-risk individuals (e.g., the chronically ill). Such compensations are necessary since insurers are not allowed to risk-rate their premiums. Without accurate risk equalization, insurers are confronted with incentives for risk selection [1]. From economical and societal perspectives, risk selection is undesirable since it may reduce: the quality of healthcare, because insurers have disincentives to improve the quality of care for subgroups for which they are undercompensated; the efficiency of care, because risk selection may be a more effective strategy for insurers to reduce costs than improving the efficiency of care; and the solidarity between low-risk and high-risk individuals when – due to market segmentation – these groups concentrate in different health plans with different premiums.

In case of imperfect risk equalization, regulators can apply three strategies to reduce incentives for risk selection: improving the risk equalization model; allowing insurers to risk rate their premiums; and establishing risk sharing between insurers and the regulator (i.e., providing insurers with retrospective compensations based on their actual costs). The first strategy is preferred since the other two have serious drawbacks. A disadvantage of risk rating is that it reduces solidarity between low-risk and high-risk individuals. Moreover, premiums may become unaffordable for subgroups of people that are seriously undercompensated by the risk equalization model. A drawback of risk sharing is that it reduces incentives for efficiency [2]. As argued by Van de Ven and Schut, decisions on risk rating and risk sharing imply complex tradeoffs between risk selection, efficiency, solidarity and affordability [3]. The only escape from this tradeoff is to improve the risk equalization model. History has shown, however, that improvements in risk equalization are not achieved overnight. In The Netherlands, it took more than 10 years to develop the risk equalization model from a simple demographic model (only compensating for age and gender) to a sophisticated health-based model (also compensating for health status). Risk rating and/or risk sharing may be necessary as long as the quality of the risk equalization model is...
considered insufficient. For deciding on the optimal levels of risk sharing and risk rating, policymakers must know the predictive accuracy of the risk equalization model, that is, the extent to which the model compensates for predictable variation in individual healthcare expenses.

Against this background, the goal of this study is to evaluate the improvement in predictive accuracy of the Dutch risk equalization model since its introduction in 1993. The outcomes are not just relevant for The Netherlands, but also for – the increasing number of – other countries relying on risk equalization in a competitive social health insurance market, since the evolution of the Dutch model is exemplary for that of risk equalization models worldwide. Our empirical evaluation primarily focuses on predictive accuracy. Note that this is not the only relevant performance aspect of risk equalization models. Other relevant aspects are, for instance, appropriateness of incentives (e.g., risk equalization models should not provide incentives for upcoding) and fairness (e.g., risk equalization models should not compensate for variation in healthcare expenses due to differences in efficiency across health plans) [1]. We will discuss these aspects later in this article in the section ‘Expert commentary’.

Evaluating the predictive accuracy of risk equalization models is a major (data-demanding) challenge. A common method is to examine the under-/overcompensation for subgroups of people in relatively poor health (and/or subgroups of people in relatively good health) that are not explicitly compensated for by these models. The databases used for estimating risk equalization models, however, hardly contain information to define such subgroups. Some studies have overcome this problem by enriching administrative data with health survey information [4-9]. In this paper, we follow a similar approach by merging administrative data from 2009 – covering (nearly) the entire Dutch population – with health survey information from a (nearly) random sample of the Dutch population in 2008. This provides us with a unique instrument for calculating the under-/overcompensations by the risk equalization models for a wide range of subgroups in the population. Since subgroups are based on information from the previous year (i.e., 2008) compared with the year of risk equalization (i.e., 2009), we expect that under-/overcompensations will be systematic.

Risk equalization in The Netherlands
The Dutch risk equalization model was implemented in 1993. The essence of this model is that insurers receive a prospective payment for each enrollee on their list, depending on particular risk characteristics of that enrollee. Since 1993, the following risk characteristics have been successively added to the model: age interacted with gender (1993), region (1995), source of income interacted with age (1999), pharmacy-based cost groups (PCGs, 2002), diagnoses-based cost groups (DCGs, 2004), socioeconomic status interacted with age (2008) and multiple-year high costs (2012). All these risk adjusters have been carefully developed in research programs initiated by the Dutch Ministry of health. In this paper, we take these risk adjusters as given and will not elaborate on the pros and cons of their particular design. Hence, we briefly describe how these characteristics are translated into 126 risk classes (in the risk equalization model of 2012).

The model includes 40 classes for age and gender, that is, 20 classes for men and 20 classes for women. The age classes are: 0, 1–4, 5–9, 10–14, 15–17, 18–24 years, 5-year cohorts up to the age of 90 and finally a class for people of 90 years or older. Information on age and gender is obtained from the insurers’ administrative databases.

In addition, the model includes 10 clusters of regions. Regions are distinguished by the four digits of the zip code, which represents a village or town or parts of either. The clustering of these zip codes is based on additional information that is not available at the individual but at the regional level (e.g., proportion of non-Western immigrants, proportion of single-households, degree of urbanization and distance to healthcare providers).

The model also includes 17 classes for source of income, interacted with age. The following four sources of income are distinguished: self-employment, disability benefits, social security benefits and other (including employment). Each of these groups is interacted with four age groups (18–34, 34–44, 45–54 and 55–64 years). Enrollees in the age groups 0–17 and ≥65 years are classified in one separate class. The information on source of income comes from the tax collector and the registration service for social benefits and assistance.

As a direct proxy for health status, the model includes 26 PCGs. The essence of PCGs is that enrollees are classified into clinically homogenous groups, based on their prior use of pharmaceuticals. Enrollees are categorized in one or more of 25 PCGs if they received at least 180 defined daily dosages (DDDs) of a certain pharmaceutical in the preceding year. For example, those who used at least 180 DDDs for insulin are categorized in the PCG for diabetes. Those without one or more of the 25 PCGs are categorized in a separate PCG. (For more information and technical details about the Dutch PCGs see [10]). The pharmaceutical information comes from a national database on drug prescriptions.

As another direct proxy for health status, the model includes 14 DCGs. The essence of DCGs is that individuals are classified into more or less homogeneous cost groups, based on a limited set of hospital (mostly inpatient) diagnoses from the previous year. Enrollees with multiple diagnoses are classified in only one DCG (that with the highest follow-up costs). Those without any of these diagnoses are categorized in a separate DCG (for more information and technical details about the Dutch DCGs, see [11,12]). The diagnostic information comes from a national database on hospital care.

Furthermore, the model includes 12 classes for socioeconomic status (SES), interacted with age. This classification is based on income, number of household members and age. For each enrollee, the income level is calculated as the household income divided by the number of household members. Those in the bottom 30% of the income distribution are categorized in SES class 1, those in the middle 30–70% are classified in SES class 2 and those in the top 30% are categorized in SES class 3. Enrollees living in a household with more than
successively applied – on the administrative data and calculate the predictive accuracy of the Dutch risk equalization model since 1993: i) estimate the risk equalization models – that have been 

wide range of subgroups in the Dutch population. (previous versions of) the Dutch risk equalization model for a datasets allows calculation of the under-/overcompensations by the two datasets for this study. The combination of these two (i.e., the owners of the administrative data) to use and to merge the Dutch Ministry of Health and the Board of Health Insurers obtained permission from CBS (i.e., owner of the survey results), risk characteristics included in the administrative data). We representative for the entire population in terms of the costs and individuals actually filled out the questionnaire (in step 2 of our anal-

tion. In total, a random group of 13,660 individuals was invited to take part in the survey of 2008. Eventually, 8735 of these indi-

cators that those with MHCs probably suffer from a chronic disease. The following six classes are distinguished: 3× costs in the top-1.5%, 3× costs in the top-4% (but not 3× in the top-1.5%), 3× costs in the top-7% (but not 3× in the top-4%), 3× costs in the top-10% (but not 3× in the top-7%), 3× costs in the top-15% (but not 3× in the top-10%) and 2× costs in the top-15% in the two preceding years (but not 3× in the top-15% in the three preceding years). Those who are not classified in one of these six groups are categorized in a separate class. The information on prior costs comes from the insurers’ administrative databases.

Data & method

For the empirical analyses, we merged administrative data with health survey information using an anonymous identification key. The administrative data are from 2009 and include individual-level information on annual medical costs and risk characteristics for almost the entire Dutch population (n = 15.6 million). This information has been used in practice for estimating the risk equalization model of 2012 and comes from various administrative sources, including insurers, the tax collector and the registration service for social benefits.

The health survey data come from a representative sample of the Dutch population in 2008. This survey is held on an annual basis by Statistics Netherlands (CBS) and includes questions on general health status, impairments, particular diseases and prior utilization of medical care. The primary goal of the survey is to monitor developments in health patterns in the Dutch population. In total, a random group of 13,660 individuals was invited to take part in the survey of 2008. Eventually, 8735 of these individuals actually filled out the questionnaire (in step 2 of our analysis we will examine the extent to which the respondents are representative for the entire population in terms of the costs and risk characteristics included in the administrative data). We obtained permission from CBS (i.e., owner of the survey results), the Dutch Ministry of Health and the Board of Health Insurers (i.e., the owners of the administrative data) to use and to merge the two datasets for this study. The combination of these two datasets allows calculation of the under-/overcompensations by (previous versions of) the Dutch risk equalization model for a wide range of subgroups in the Dutch population.

We used a four-step procedure to evaluate the improvement in predictive accuracy of the Dutch risk equalization model since 1993: i) estimate the risk equalization models – that have been successively applied – on the administrative data and calculate the predicted costs per model per individual, ii) merge the predicted and actual annual costs (in 2009) with the health survey data (for each individual in the survey), iii) identify subgroups of people in poor health (in the survey) and iv) calculate the average under- or overcompensation per model per subgroup.

Step 1: estimate eight risk equalization models & calculate the predicted costs

In order to evaluate the improvement of the risk equalization model since 1993, we estimated the eight risk equalization models presented in Table 1. These models approximate the actual models that have been successively applied in The Netherlands in the period 1993–2012. Similar to the estimation procedure used in practice, we estimated these models by the least-squares method with medical expenses in 2009 as the dependent variable and the risk classes described in the previous section as independent dummy variables. In terms of R-squared, the most important risk adjusters in the risk equalization model are age/gender, PCGs, DCGs and multiple-year high costs. Model 7 includes all risk adjusters (i.e., 126 risk classes) and thereby represents the Dutch risk equalization model of 2012. The R-squared of this model is nearly 30%. Although this R-squared indicates high predictive power, it does not tell whether this model ‘sufficiently’ compensates for predictable variation in medical expenses. One of the reasons is that we do not know how high the R-squared would have been if the model fully corrected for all predictable variation in individual annual medical expenses. Therefore, a better way to indicate the performance of risk equalization models is to examine the under-/overcompensations for subgroups that are not explicitly included in these models. However, since most administrative databases used for estimating risk equalization models hardly contain additional health information that has not (yet) been included in these models, defining such subgroups is a major challenge. In this study, we overcome this problem by enriching the administrative data with health survey information.

Table 1. Eight models regressing healthcare expenditures in 2009 on various sets of risk adjusters using administrative data (n = 15.6 million).

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>R-squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>Constant (no risk adjusters)</td>
<td>0.00%</td>
</tr>
<tr>
<td>1</td>
<td>Model 0 + 40 classes for age/gender</td>
<td>5.97%</td>
</tr>
<tr>
<td>2</td>
<td>Model 1 + 10 clusters for region</td>
<td>6.01%</td>
</tr>
<tr>
<td>3</td>
<td>Model 2 + 17 classes for source of income</td>
<td>6.83%</td>
</tr>
<tr>
<td>4</td>
<td>Model 3 + 26 PCGs</td>
<td>15.92%</td>
</tr>
<tr>
<td>5</td>
<td>Model 4 + 14 DCGs</td>
<td>24.99%</td>
</tr>
<tr>
<td>6</td>
<td>Model 5 + 12 classes for SES</td>
<td>25.04%</td>
</tr>
<tr>
<td>7</td>
<td>Model 6 + 7 classes for MHC</td>
<td>29.61%</td>
</tr>
</tbody>
</table>

DCGs: Diagnoses-based cost groups; MHC: Multiple-year high cost; PCGs: Pharmacy-based cost groups; SES: Socioeconomic status.
Table 2. Costs and characteristics in 2009 in the administrative database (of 2009) and in the survey sample (of 2008) that successfully merged with the administrative data (of 2009).

<table>
<thead>
<tr>
<th>Costs and characteristics in 2009</th>
<th>Administrative data</th>
<th>Survey sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enrollees in 2009 (n)</td>
<td>15,588,677</td>
<td>8735</td>
</tr>
<tr>
<td>Mean actual costs 2009*</td>
<td>1570</td>
<td>1593</td>
</tr>
<tr>
<td>Mean predicted costs 2009</td>
<td>1570</td>
<td>1535</td>
</tr>
<tr>
<td>Mean undercompensation* by model 7†</td>
<td>0</td>
<td>58</td>
</tr>
<tr>
<td>Men, 0–24 years</td>
<td>15.1%</td>
<td>16.5%</td>
</tr>
<tr>
<td>Women, 0–24 years</td>
<td>14.5%</td>
<td>15.6%</td>
</tr>
<tr>
<td>Women, 25–49 years</td>
<td>17.4%</td>
<td>16.3%</td>
</tr>
<tr>
<td>Women, 50–74 years</td>
<td>14.6%</td>
<td>15.5%</td>
</tr>
<tr>
<td>Women, 75 years and older</td>
<td>2.7%</td>
<td>2.7%</td>
</tr>
<tr>
<td>Classified in at least one PCG</td>
<td>4.5%</td>
<td>4.3%</td>
</tr>
<tr>
<td>Classified in a DCG</td>
<td>2.5%</td>
<td>2.4%</td>
</tr>
<tr>
<td>Classified in a MHC group</td>
<td>7.3%</td>
<td>7.2%</td>
</tr>
</tbody>
</table>

\*2009 Euros.
\*Mean undercompensation = mean actual costs - mean predicted costs.
PCGs: Pharmacy-based cost groups.
DCGs: Diagnoses-based cost groups; MHC: Multiple-year high cost.

Although these deviations by themselves are not statistically different from zero (p > 0.05), a potential problem is that they may influence whether or not under-/overcompensations for specific subgroups are statistically different from zero. To overcome this potential problem, we multiplied the predicted costs by a factor $F_m$ (where $F_m$ equals the average actual costs in 2009 divided by the average predicted costs for 2009 according to model $m$) such that per model the average actual costs in the survey sample as a whole equal the average predicted costs.

**Step 3: identify subgroups of people in poor health**

Based on the survey data we defined several subgroups of high-risk individuals. In order to mitigate the effect of random variation (due to the limited sample size of the survey), we focused on subgroups containing at least 400 individuals. Eventually, we were able to define the following subgroups based on self-reported health:

- at least one self-reported chronic condition;
- worst score on physical health status according to the SF-12 summary score (SF-12 is a short version of the SF-36 health survey including 12 questions on relevant components of physical and mental health from which summary scores for physical health and mental health can be construed [14]);
- worst score on mental health status according to the SF-12 summary score;
- health status is regarded as ‘bad, poor or not so good’;
- mobility impairment according to the OECD mobility indicator (the OECD mobility indicator is based on three questions about mobility [15]);
- arthritis of hips or knees in the last 12 months;
- serious back problems in the last 12 months;
- serious neck problems in the last 12 months;
- obese (BMI >30).

In addition, we defined the following subgroups based on (self-reported) prior utilization of healthcare:

- contact with a general practitioner in the last 12 months;
- contact with a medical specialist in the last 12 months;
- hospitalization in the last 12 months;
- use of prescribed drugs in the last 14 days;
- use of durable medical equipment in the last 12 months;
- use of physiotherapy in the last 12 months.

**Step 4: calculate the average under- or overcompensation per model per subgroup**

As a last step, we calculated the average under-/overcompensation for the 18 subgroups defined in step 3. Since the subgroups are based on information from the previous year, under- or overcompensations implies that a model leaves incentives for risk selection. The results will be presented in...
terms of average undercompensation which is defined as the average actual expenses in 2009 for a certain subgroup minus the average predicted expenses in 2009 for that subgroup. Note that a negative undercompensation would imply overcompensation.

We expect that under-/overcompensations will – to a major extent – be systematic. In other words, similar results are expected when the same subgroups would have been based on information from 2009 and the same risk equalization models would have been estimated on information from 2010. We will therefore formulate our results in a general way, that is, “the under-/overcompensation in year t of subgroups based on information from year t-1”.

Results

Table 3 shows the average undercompensation per person in year t per model and per subgroup based on self-reported health in year t-1. Five interesting results can be observed. In the first place, it appears that under model 0 (in which the predicted individual costs for year t simply equal the population average costs for year t), all 18 subgroups are (substantially) undercompensated. This would provide competing insurers with substantial incentives for risk selection. For example, people who reported at least one chronic condition (32.6% of the population) in a certain year are undercompensated in the following year by an average €1,466 per person (i.e., 93% of the population average costs per person).

In the second place, Table 3 shows that undercompensations substantially decrease as more relevant risk adjusters are added to the model. To highlight this pattern, Figure 1 presents the undercompensation of five subgroups graphically. In general terms, the largest improvements result from the introduction of age/gender, source of income, PCGs, DCGs and MHC. The impact of region and SES is relatively minor. A noteworthy finding is that the addition of the risk adjuster “source of income” (model 3 compared with model 2) results in a substantial reduction of the average undercompensation per person in the five subgroups in Figure 1, but results only in a modest increase of the R-squared in Table 1. Compared with model 0 (without risk adjusters), model 7 (including all seven risk adjusters) reduces the total undercompensation for the 18 subgroups in Table 3 by 47% (people that suffered from migraine or serious headaches in the last 12 months) to 90% (people who suffer from obesity). For the subgroup of people who reported at least one chronic condition the reduction in undercompensation by model 7 equals 71%.

In the third place, Table 3 shows that even model 7 (including all seven risk adjusters) still undercompensates people in these subgroups by on average €77 to €858 per person per year. The undercompensation is statistically significantly different from zero for the following subgroups: people who reported at least one chronic condition (who are on average undercompensated by €426 per person per year), people with the worst score on physical health (€754), people with the worst score on mental health (€479), people who evaluate their health status as bad, poor or moderate (€646), people who suffered from hypertension (€413) in the last 12 months, people who visited a medical specialist in the last 12 months (€333), people who used prescribed drugs in the last 14 days (€270) and people who used physiotherapy in the past 12 months (€358). These significant undercompensations imply that even the current model (2012) – with sophisticated pharmacy-, diagnoses- and prior-cost-based risk adjusters – still leaves substantial incentives for risk selection. Note that – since the risk equalization model is based on a zero-sum principle – the undercompensations in Table 3 imply overcompensations for the complementary groups. For example, the average undercompensation of €426 (model 7) for people with a chronic condition in year t-1 implies an average overcompensation of €206 (i.e., €426 × [32.6%/67.4%]) for people without a chronic condition in year t-1. For all groups in Table 3, insurers have incentives to selectively attract the complementary group, even with model 7.

A fourth relevant result concerns the people suffering from asthma, chronic bronchitis, lung emphysema or COPD. Although these health problems are included in the Dutch PCG-classification – which explains the substantial decrease in average undercompensation by model 4 – models 4–7 do not eliminate the undercompensation for this subgroup completely. The explanation may be that not all of these people are included in the relevant PCG because they do not use more than 180 DDDS of the pharmaceuticals on which this PCG is based. We expect similar results for other diseases included in the Dutch PCG-classification. This implies that incentives for risk selection are not only present for health problems that have not (yet) been included in the PCG-classification, but also for health problems that have been included.

In the fifth place, it can be observed that the undercompensation for a subgroup does not necessarily decrease as more risk adjusters are added to the model. The introduction of DCGs (going from model 4 to model 5), for instance, results in an increase of the average undercompensation for people who suffered from serious neck problems in the last 12 months by €100 per person. Apparently, these people have a relatively low probability of being classified in a DCG. Since the introduction of DCGs will in general result in a decrease of the coefficients for the other risk adjusters (because of correlation between risk adjusters and because the risk equalization system is based on the zero sum principle), the compensation for this particular subgroup slightly decreases (and consequently the undercompensation slightly increases). So, it is possible that the introduction of new risk adjusters increases the under-/overcompensation for specific subgroups, particularly when the new risk adjusters are correlated with the risk adjusters that are already included in the risk equalization model.

Directions for improvement

Table 3 does not only indicate the imperfections in the Dutch risk equalization model but also points at possibilities for...
Table 3. The average undercompensation per person (in 2009 euros) in year t per model per subgroup based on survey information from year t-1.

<table>
<thead>
<tr>
<th>Subgroups based on information from year t-1</th>
<th>Size of subgroup</th>
<th>Average undercompensation in year t (average actual costs minus average predicted costs)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Model 0 (constant)</td>
</tr>
<tr>
<td>At least one chronic condition</td>
<td>32.6%</td>
<td>1466**</td>
</tr>
<tr>
<td>Worst score physical health (SF-12)</td>
<td>9.3%</td>
<td>3067**</td>
</tr>
<tr>
<td>Worst score mental health (SF-12)</td>
<td>9.3%</td>
<td>1148**</td>
</tr>
<tr>
<td>Health status regarded as bad, poor or moderate</td>
<td>19.6%</td>
<td>2275**</td>
</tr>
<tr>
<td>Mobility impairment (OECD)</td>
<td>6.3%</td>
<td>3861**</td>
</tr>
<tr>
<td>Migraine or serious headaches in last 12 months</td>
<td>10.8%</td>
<td>203</td>
</tr>
<tr>
<td>Hypertension in last 12 months</td>
<td>12.4%</td>
<td>1886**</td>
</tr>
<tr>
<td>Asthma, chronic bronchitis, lung emphysema or COPD in last 12 months</td>
<td>6.1%</td>
<td>1742**</td>
</tr>
<tr>
<td>Arthritis in hips or knees in last 12 months</td>
<td>10.3%</td>
<td>2025**</td>
</tr>
<tr>
<td>Serious back problems in last 12 months</td>
<td>8.4%</td>
<td>1084**</td>
</tr>
<tr>
<td>Serious neck problems in last 12 months</td>
<td>7.9%</td>
<td>953**</td>
</tr>
<tr>
<td>Obesity (BMI &gt;30)</td>
<td>9.1%</td>
<td>896**</td>
</tr>
<tr>
<td>Contact with general practitioner in last 12 months</td>
<td>73.3%</td>
<td>257**</td>
</tr>
<tr>
<td>Contact with medical specialist in last 12 months</td>
<td>41.2%</td>
<td>1063**</td>
</tr>
<tr>
<td>Hospitalization in last 12 months</td>
<td>6.8%</td>
<td>2566**</td>
</tr>
<tr>
<td>Use of prescribed drugs in last 14 days</td>
<td>39.7%</td>
<td>1288**</td>
</tr>
<tr>
<td>Regular use of durable medical equipment in last 12 months</td>
<td>5.6%</td>
<td>3260**</td>
</tr>
<tr>
<td>Use of physiotherapy in last 12 months</td>
<td>19.5%</td>
<td>904**</td>
</tr>
</tbody>
</table>

1 For composition (and underlying questions) of this score, see [14].
2 For composition (and underlying questions) of this variable, see [15].
3 Regular use of at least one of the following categories: aids for walking (e.g., walker), wheelchair (hand or electronic), orthopedic shoe, prosthesis (arm or leg), splint, incontinence materials, catheter and stoma devices.
*Statistically significantly different from zero with p < 0.05;
**Statistically significantly different from zero with p < 0.01.
DCGs: Diagnoses-based cost groups; MHC: Multiple-year high cost; PCGs: Pharmacy-based cost groups; SES: Socioeconomic status.
further improvement. In general, the quality of the model will substantially improve if these 18 subgroups would be included as additional risk adjusters. One way to do this would be to use the survey results to identify these groups. A problem however is that these data are not available for the entire population. In addition, it would be very costly – and practically impossible – to hold the survey among the entire population. A better alternative may be to rely on administrative data to identify these groups.

At this moment, all-inclusive information on drug prescriptions and hospital treatments is already available. However, the undercompensations for people who used prescribed drugs in the last 14 days (€270), people who were hospitalized in the last 12 months (€446) and people who visited a medical specialist in the last 12 months (€333) indicate that the PCGs and DCGs in the current risk equalization model do not fully exploit this information. Therefore, improvement of the PCGs and DCGs may further reduce the undercompensations in Table 3. Furthermore, the average undercompensation for people who regularly used durable medical equipment in the last 12 months (€858) indicates that a health classification based on the prior use of durable medical equipment may be a fruitful direction for further improvement. This is in line with the results of Van Kleef and Van Vliet [16]. Finally, the undercompensation for people who used physiotherapy in the last 12 months (€358) indicates that a classification based on physiotherapy may be another improvement. Similar to drug prescription and hospital treatment, information on the use of durable medical equipment and physiotherapy is already available in administrative databases.

Note that inclusion of more and better morbidity-based risk adjusters may reduce the impact of other risk adjusters, particularly that of the most indirect measures of health status such as socioeconomic status and region. Since the impact of these risk adjusters is already small (Tables 1 & 3), the role of these variables as health adjusters may become redundant.

**International perspective**

Risk equalization is not only applied in The Netherlands, but also in other countries with regulated health insurance markets. Outside The Netherlands, the most sophisticated prospective models can be found in Germany, Belgium and the USA. An interesting question is how the Dutch model is performing compared with these other models. Quantitative comparison, however, is rather difficult since outcomes in terms of predictive accuracy do not only depend on the design of the risk equalization models but also on patterns in the underlying databases on which these models are evaluated. Thus, the observation that this study finds an R-squared of 29.6% for the Dutch model while Buchner et al. find an R-squared of 23.9% for the German model ([17], page 258) and Winkelman and Mehmud find R-squares ranging from 18.4 to 27.4% for several common models in the USA ([18], fifth column of Table IV.6 on page 17) is only a rough indication that the predictive accuracy of the Dutch model is relatively high. For an unbiased comparison of the different risk equalization models they should be evaluated on the same database. Such an analysis is outside the scope of this paper.

A simple comparison of the structure of the Dutch risk equalization model with other models, however, reveals at least two potential improvements for the Dutch model. The first concerns the source and amount of diagnostic information used for DCGs. The DCG-HCC model [17] in Germany as well as the CMS/HCC-model [19], the CDPS-model [20], ACG-model [21] and the CRG-model [22] are so-called all-encounter models, which means that all data sources containing diagnostic information are taken into account. This is different in the Dutch model of 2012 where DCGs are solely based on inpatient information. Therefore, a potential improvement for the Dutch model would be to include diagnostic information from ambulatory services as well. In 2013, the Dutch government made the DCGs also dependent on diagnostic information from prior outpatient visits in hospitals. Since diagnostic information from general practitioner visits is not yet available in the current administrative databases used for risk equalization, further extension would require additional data collection.

The second potential improvement concerns the structure of DCGs. In the German and US models mentioned above, diagnoses are grouped into so-called disease hierarchies. These hierarchies allow classifying individuals with multiple (unrelated) diagnoses into multiple risk classes. This is different in the Dutch model where diagnoses are grouped in DCGs just by
similarities in follow-up costs and individuals can be classified in one DCG only. The latter is necessary to avoid overcompensation of individuals with multiple diagnoses related to the same disease but grouped in different DCGs. Replacing DCGs with a system of disease hierarchies may improve the predictive accuracy for individuals with comorbidity. At this moment, however, the possibilities for disease hierarchies in The Netherlands are limited. The reason is that the current DCGs are not directly based on diagnostic codes, but on a classification of healthcare products containing only limited diagnostic information. Replacing the current DCGs with a system of disease hierarchies would therefore require additional data collection. A potential improvement with the currently available data, however, is to allow for a restricted set of multiple DCGs and/or interactions among DCGs.

Conclusion
This paper has quantified the improvement in predictive accuracy of the Dutch risk equalization model since 1993. Based on administrative data from 2009 and survey information from 2008, we find that the model has substantially improved: for people who reported a chronic condition in year t-1 (33% of the population) the current risk equalization model (2012) reduces the average undercompensation in year t by 71%. But even this model – including sophisticated pharmacy-, diagnoses- and prior-cost-based risk adjusters – still undercompensates for all 18 subgroups identified in our analysis. For the subgroup of people who reported a chronic condition in year t-1, the current risk equalization model leads to an average undercompensation in year t of more than €400 per person. Such undercompensation leaves insurers with substantial incentives for risk selection, which may threaten quality of care and solidarity between low-risk individuals (e.g., the young and healthy) and high-risk individuals (e.g., the elderly and chronically ill). The good news, however, is that with the currently available data the model can probably be improved by i) extending PCG- and DCG-classifications, ii) allowing for a restricted set of multiple DCGs and/or interactions among DCGs, iii) a health classification based on the prior use of durable medical equipment and iv) a health classification based on the prior use of physiotherapy.

Expert commentary
Our findings may raise the following two questions:

- Can risk equalization ever compensate for all predictable variation in individual medical expenses?
- Should risk equalization compensate for all predictable variation in individual annual medical expenses?

With respect to the first question it is important to note that ‘predictive accuracy’ is not the only performance aspect of risk equalization models. These models should also meet other criteria like fairness, feasibility and appropriateness of incentives [1]. The latter, for instance, is an important motive for applying DDD-thresholds to PCGs. Without such thresholds insurers may be confronted with substantial perverse incentives to provide enrollees with more prescribed drugs in order to receive higher risk equalization payments in later years. However, such restrictions do not only reduce perverse incentives, but (on average) also result in undercompensation for people in poor health since fewer of these people will be classified in a PCG. The same holds for restrictions on DCGs. In the absence of new risk adjusters (that meet the above-mentioned criteria) alternative measures may be necessary to reduce incentives for risk selection. One alternative may be to overcompensate PCGs and DCGs such that the undercompensation for the subgroups of people in poor health (including those not classified in a PCG or DCG) as such in Table 3 are largely eliminated. A drawback, however, may be that this increases incentives for risk selection within the group of people in poor health. This drawback can be counteracted by a further refinement of the relevant risk adjusters. Another alternative may be to allow (some degree) of risk rating. This strategy will not only reduce incentives for risk selection but will also reveal the information surplus of insurers relative to the current risk equalization model. The latter may indicate what (type of) information is missing in the current model. A drawback of risk rating, however, is that solidarity between low-risk and high-risk individuals will decrease and that health-plan premiums may become unaffordable for subgroups that are substantially undercompensated. This drawback can be counteracted by direct subsidies to the relevant households, for example, by adjusting the existing income-related subsidies that more than 4 million households currently receive (‘zorgtoeslag’). Another way to reduce remaining incentives for risk selection is risk sharing [13,23,24]. As mentioned earlier, however, risk sharing will not only reduce the insurers’ incentives for risk selection but also their incentives for efficiency [2]. So, as long as the risk equalization model does not sufficiently compensate for predictable variation in individual health expenses, regulators are confronted with tradeoffs between risk selection, efficiency and solidarity.

With respect to the second question it is important to note that society/regulators may not want to compensate for all predictable variation in individual medical expenses [25,26]. Schokkaert et al. distinguish between C- and R-type cost variation [27]. The first refers to cost variation for which society wants compensation (typically cost variation related to age, gender and health). The second refers to variation for which society does not want compensation (typically cost variation related to supplier characteristics). However, if the regulator decides not to compensate for R-type cost variation, it is consistent – for reasons of fairness and incentives – to allow insurers to risk rate their premiums according to this type of cost variation.

Five-year view
The incentives for Dutch insurers to engage in risk selection do not only depend on the quality of the risk equalization model
The introduction of risk equalization in 1993 came with substantial risk sharing: for each euro profit/loss on average 97 cents were shared with the regulator, that is, insurers were liable for only 3 cents. However, as the risk equalization model was improved, the financial risk for insurers was increased (Figure 2). From 2011 to 2012, the average financial risk increased from 74 to 92%. The government has announced another increase for the near future: per 2015, insurers should bear the full financial risk. A positive effect of more financial risk is that incentives for efficiency will increase; a negative effect, however, is that incentives for risk selection will increase as well (since insurers will be liable for a larger share of the undercompensation of certain subgroups).

In order to stimulate insurers to invest in efficiency and not in risk selection, the government acknowledges the necessity of further improving the risk equalization model. Per classification – which was mainly based on inpatient hospital information – is extended with outpatient hospital information. A recent study carried out by Van Kleef et al. indicates that this extension reduces the average undercompensation of people who reported a chronic condition in year t-1 from €426 (Table 3) to €381 per person per year [28]. In addition, the government plans to introduce a new risk adjuster based on the use of durable medical equipment by 2014. In order to evaluate the effects of new risk adjusters, we recommend monitoring undercompensation of relevant subgroups very carefully. The set of subgroups in Table 3 may serve as a starting point. For a more comprehensive evaluation, this set could be extended with other relevant subgroups of people in relatively poor health (and/or subgroups of people in relatively good health) that are not explicitly compensated for by the risk equalization model. Such subgroups may be identified by using prior costs, prior utilization and diagnoses from survey information as well as administrative data. In addition, it may be useful to apply other quantitative measures than under-/overcompensation of subgroups.

Policymakers should seriously take into account that despite the above-mentioned improvements of the Dutch risk equalization model some incentives for risk selection will remain. This means there is no guarantee that insurers will not engage in risk selection. We therefore recommend the regulator to monitor the insurers’ behavior very carefully. Any signal of risk selection may be a motive for recalling risk sharing and/or allowing some degree of risk rating. To measure (signals of) risk selection it is crucial to have an overview of the possible forms of risk selection as well as their potential effects. In another contribution to this issue of Expert Review of Pharmacoeconomics & Outcomes Research, we provide such an overview.

Acknowledgements
The authors thank the following persons for their valuable comments on previous versions of this article: three anonymous reviewers, the members of the Risk Adjustment Network (RAN), in particular Florian Buchner, and the Advisory Group (‘Begeleidingsgroep’). 

Disclaimer
The opinions and views expressed in this paper are those of the authors and do not necessarily reflect the opinions or views of the NPCF or those of the persons mentioned above. The responsibility for the content of this article fully rests with the authors.

Financial & competing interests disclosure
The authors gratefully acknowledge the Dutch Federation of Patients/Consumers (NPCF) for financing this study in part. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.
Key issues

- The Netherlands relies on risk equalization to compensate competing insurers for predictable variation in individual medical expenses.
- The predictive accuracy of the Dutch risk equalization model substantially improved over the years, for example, for people who reported a chronic condition in year t-1 the current risk equalization model (2012) reduces the average undercompensation in year t by about 70% compared with no risk equalization at all.
- Even the current model, however, undercompensates for particular subgroups in the population, for example, for people who reported a chronic condition in year t-1 it leads to an average undercompensation in year t of more than €400 per person per year.
- Since Dutch insurers are not allowed to risk rate their premiums, undercompensation of identifiable subgroups confronts them with incentives for risk selection, which threatens quality of care as well as solidarity between low-risk and high-risk individuals.
- Further improvements in risk equalization are necessary to safeguard public objectives.
- Promising potential improvements with the currently available data are i) extending pharmacy-based cost groups (PCGs) and/or diagnoses-based cost groups (DCGs), ii) allowing for a restricted set of multiple DCGs and/or interactions among DCGs, iii) developing a health classification based on the prior use of durable medical equipment and iv) developing a health classification based on the prior use of physiotherapy.
- As long as the risk equalization model is imperfect, it may be desirable to reduce remaining incentives for risk selection by risk sharing and/or allowing some degree of risk rating.
- In case of imperfect risk equalization, decisions on risk sharing and risk rating require complex tradeoffs between risk selection, efficiency and solidarity.

References

Papers of special note have been highlighted as:
• of interest
** of considerable interest
** Provides a conceptual framework on risk equalization (i.e., risk adjustment) in competitive health insurance markets.
** Discusses the trade-off between efficiency and selection when paying competing health insurers.
** Discusses the trade-off between risk selection, efficiency, solidarity and affordability in competitive health insurance markets with imperfect risk equalization.
** Examines the potential improvement of durable medical equipment as a risk adjuster.
** Describes recent developments regarding risk equalization in Germany.
19 Pope GC, Kautter J, Ellis RP et al. Risk adjustment of Medicare capitation payments


• Discusses different options of risk sharing as a supplement to imperfect risk equalization.


