Cost Variation in Diabetes Care across Dutch Care Groups?

Sigrid M. Mohnen, Claudia C. Molema, Wouter Steenbeek, Michael J. van den Berg, Simone R. de Bruin, Caroline A. Baan, and Jeroen N. Struijs

Objective. The introduction of bundled payment for diabetes care in the Netherlands led to the origination of care groups. This study explored to what extent variation in health care costs per patient can be attributed to the performance of care groups. Furthermore, the commonly applied simple mean aggregation was compared with the more advanced generalized linear mixed model (GLMM) to benchmark health care costs per patient between care groups.

Data Source. Dutch 2009 nationwide insurance claims data of diabetes type 2 patients (104,544 patients, 50 care groups).

Study Design. Both a simple mean aggregation and a GLMM approach was applied to rank care groups, using two different health care costs variables: total treatment health care costs and diabetes-specific specialist care costs per diabetes patient.

Principal Findings. Care groups varied slightly in the first and mainly in the second indicator. Care group variation was not explained by composition. Although the ranking methods were correlated, some care groups' rank positions differed, with consequences on the top-10 and the low-10 positions.

Conclusions. Differences between care groups exist when an appropriate indicator and a sophisticated aggregation technique is used. Currently applied benchmarking may have unfair consequences for some care groups.

Key Words. Care group, variation, health care costs, GLMM, bundled payment, Vektis

Over the last years, several countries introduced payment reforms in chronic care to optimize the performance of their health care systems (Van Herck et al. 2010; De Bruin, Baan, and Struijs 2011; Doran et al. 2011; Song et al. 2011; Tsiachristas et al. 2013). In the Netherlands, a bundled payment (BP) model was introduced in 2007, designed to facilitate multidisciplinary cooperation between care providers by eliminating financial barriers between care sectors

and disciplines (Struijs and Baan 2011). Under the BP model, health insurers are able to purchase the different care components of disease management programs (DMPs) as a single package from so-called care groups. Care groups are provider-led organizations that serve as principal contractors of BP contracts and are responsible for coordinating and ensuring the delivery of services included in the DMP for diabetes patients.

Care groups are financially and clinically responsible for all patients enrolled in the DMP (Struijs and Baan 2011; De Bakker et al. 2012). In 2009 there were about 50 and in 2011 already 100 care groups in the Netherlands, all of which had a BP contract for diabetes care, covering virtually the entire country. On average, about 80 general practitioners (GPs) participate in a care group (De Jong-van Til et al. 2012). Care groups aim to stimulate the collaboration between care providers, which is expected to improve quality and continuity of diabetes care within primary care and between primary and secondary care. Due to the formation of care groups, it is expected that more diabetes patients will be well controlled, leading to fewer patients consuming hospital-based specialist care and eventually to a slowdown of the cost growth.

As care groups are financially and clinically accountable for the services contracted for all enrolled patients, there is an increasing interest to benchmark the performance of care groups among insurers (accountability information), the government (oversight), citizens (patient choice information), and care groups themselves (reflective information). Measuring performance correctly is a challenge public health researchers are facing in a “learning health care system” (Berenson, Pronovost, and Krumholz 2013). Until now, performance of care groups has mostly been studied by comparing simple mean aggregations of patient-level care costs, quality, or utilization measures (Achmea 2011; Zorgverzekeraars Nederland 2011; Adviesgroep Ketenzorg 2012; Van Schooten et al. 2012; Landelijke Organisatie voor Ketenzorg 2013),

Address correspondence to Sigrid M. Mohnen, Ph.D., National Institute for Public Health and the Environment (RIVM), Centre for Nutrition, Prevention, and Health Services, PO Box 1, 3720 BA Bilthoven, The Netherlands; e-mail: sigrid.mohnen@rivm.nl. Claudia C. Molema, M.Sc., Simone R. de Bruin, Ph.D., and Jeroen N. Struijs, Ph.D., are with the National Institute for Public Health and the Environment (RIVM), Centre for Nutrition, Prevention, and Health Services, Bilthoven, The Netherlands. Wouter Steenbeek, Ph.D., is with the Netherlands Institute for the Study of Crime and Law Enforcement (NSCR), Amsterdam, the Netherlands. Michael J. van den Berg, Ph.D., is with the National Institute for Public Health and the Environment (RIVM), Centre for Health and Society, Bilthoven, The Netherlands. Caroline A. Baan, Ph.D., is with the National Institute for Public Health and the Environment (RIVM), Centre for Nutrition, Prevention, and Health Services, Bilthoven, The Netherlands; Scientific Centre for Transformation in Care and Welfare (Tranzo), University of Tilburg, Tilburg, the Netherlands.

which is surprising given the long tradition of controlling for patient case mix in health care provider performance profiling (e.g., see Salem-Schatz et al. 1994; Iezzoni et al. 1996). Case mix refers to the patient characteristics that may affect health care cost, which has to be taken into account for fair comparison across health care providers.

In this study we focus on cost variation to evaluate the performance of care groups. Large cost variations that are not explained by case mix, patient's illness (kind and status of the illness), or preferences (for health care consumption) are considered to indicate a below-standard economic performance of those care groups with significantly higher costs, thereby identifying room for (further) improvement (Wennberg 2011). The aforementioned costs studies (Achmea 2011; Van Schooten et al. 2012) concluded that huge cost differences between care groups exist, and subsequently stated that large efficiency gains are potentially (still) achievable. However, both studies did not present the corresponding confidence intervals (CIs) for these point estimates. Simple mean aggregation without CIs ignores that care groups with more enrolled patients are measured more reliably than those with fewer patients. Thus, it remains unclear to what extent these differences between care groups are attributable to real performance respectively to other factors outside the power of care groups or coincidence.

In addition the multilevel problem, that is, that data are structured in multiple levels of clustered relationships between sets of units (Hox 2002), tends to be ignored which can lead to incorrect inferences. For example, in the case of care groups, patients (and their characteristics) are clustered within GPs, who in turn are clustered within care groups. As a consequence, care group variation based on simple mean aggregation may be an artifact of the composition effect (i.e., the distribution of patients or GPs) and not a real "context" effect at the care group level (Duncan, Jones, and Moon 1998; Malmström, Sundquist, and Johansson 1999). Therefore, the patient level as well as the GP level have to be accounted for when trying to investigate the isolated impact of the "context" care group on costs and the variation in costs.

Improving upon these studies, we use a generalized linear mixed model (GLMM) to capture the context effect of care groups on individual health care costs. This approach combines two long-standing methodological issues in public health research, namely the modeling of (highly skewed) health care costs (e.g., Manning and Mullahy 2001; Kilian et al. 2002) and health care provider profiling (e.g., Christiansen and Morris 1997). In short, the generalized linear model is used to model the skewed costs distribution, while random components model the nested structure of the data.

The application of a random-effects approach to study differences between care groups may reveal different results than the previous Dutch studies have found based on simple mean aggregation. The aims of our study were therefore (1) to investigate to what extent the variation of treatment costs of patients with diabetes type 2 can be attributed to the performance of care groups; and (2) to investigate if the more sophisticated approach finds significant differences between care groups and whether this approach results in a different performance ranking than the currently employed simple mean aggregation.

The analyses were performed for two indicators, namely (1) total treatment health care costs per patient, and (2) diabetes-specific specialist care costs per patient. The first indicator was chosen because the general assumption is that care groups reduce total treatment health care costs of diabetes patients. Total treatment health care costs cover all costs of primary and secondary care as well as medication (diabetes-related and nonrelated). However, to the extent that a large part of the health care utilization occurred outside the scope of the care groups, total treatment health care costs is biased by confounding variables, such as treatments for nondiabetes-related specialist care. Therefore, we repeated the analyses with diabetes-specific specialist care costs. In the Netherlands, specialist care takes place exclusively in secondary care.

METHODS

Data Sources

We used 2009 health insurance claims, which were obtained from the health care information center Vektis. Vektis collects and manages health claims data on all health care procedures covered by the Dutch basic statutory insurance package, including the costs for compulsory excesses (De Bo 2011). The Vektis databases covered 92 percent of all insured people living in the Netherlands in 2009 and comprised data of all Dutch health insurers. Vektis data also contain personal information on the policyholders, including date of birth and sex. To ensure the privacy of patients and care providers Vektis pseudonymised the patients' citizen service numbers (national identification numbers) and care provider codes.

Study Population

We analyzed health claims data of adult diabetes type 2 patients who were insured and alive in the year 2009 and were registered with GPs participating

in care groups paid on a BP basis. GPs were classified using specific procedure code numbers, which were used by care groups to claim BP fees. For this study, we selected 104,544 cases from estimated 150,000 type 2 diabetes patients enrolled in BP in 2009. Health claims data are not ready-to-use and need to undergo several checks (Appendix S1). Besides the checks, patients could be excluded based on exclusion criteria. Exclusion criteria for patients were having gestational diabetes (identified by procedure code numbers of the hospital payment system); having diabetes type 1; being younger than 18 years; missing or incorrect information on age, sex, care group, or provider; other reasons, including changing insurance company, care group, or GP during the study period. Moreover, we only included patients enrolled in a DMP for the whole year (Appendix S1). Our population is described best as adult diabetics with complete health claims data for the year 2009; not classified as gestational but defined by: (1) BP health claim codes of DMPs for type 2 diabetes patients and (2) ATC codes for noninsulin diabetes medication (A10B “blood glucose lowering drugs”).

Dependent Variables of Indicators

The variables of main interest were total treatment health care costs per diabetes patient in 2009, and diabetes-specific specialist care costs per diabetes patient in 2009. In Appendix S2A and B, descriptive statistics are presented in Euros.

The total treatment health care costs per patient were calculated by adding (1) general practice costs (i.e., consultation fees and capitation allowances, costs for practice nurses, and costs for integrated care, i.e., BPs), (2) costs of specialist care (i.e., outpatient costs, day-patient costs, and inpatient costs), and (3) other treatment costs (i.e., costs of pharmaceutical care, medical aids, physiotherapy, exercise therapy, speech therapy, occupational therapy, dietetics, patient transport, and mental health care).

The diabetes-specific specialist care costs per patient were calculated by adding outpatient costs, day-patient costs, and inpatient costs for diabetes-specific specialist care. Diabetes-specific was identified by the diagnosis “diabetes,” which is part of the hospital declaration codes.

Statistical Analysis

The dataset has a nested structure: patients are nested within GPs, which are in turn nested within care groups. First, descriptive analyses were performed

to get an overview of the characteristics of the included care groups, participating GPs, and patients enrolled in the DMPs of the care groups.

To answer the first research question regarding care groups and the treatment health care costs of patients with diabetes type 2, we studied the influence of care groups on two indicators of health care costs. To model costs with the aim of decomposing composition effects from context effects in the nested dataset, we used a GLMM with a log-normal distribution and identity link function using the GLIMMIX procedure in SAS. The linear predictor equals $\eta_{ijk} = \gamma_{000} + \Gamma X_{ijk} + \Lambda G_{ojk} + v_{00k} + u_{ojk} + \varepsilon_{ijk}$ which models the health care costs associated with patient i nested within GP j nested within health care group k ; γ_{000} is the grand mean intercept; X_{ijk} is the matrix of patient characteristics which may affect health care costs (e.g., age (centered on 67.1 years), sex (ref. female), and comorbidity (zero, one, two, three, or more)) and Γ are the patient-level parameter estimates; G_{ojk} is a matrix of predictors of GP-level characteristics (e.g., sex of GP and years of experience were no significant confounders and were therefore excluded from the models); and Λ refers to the GP-level parameter estimates. These patient-level and GP-level characteristics thus account for *composition* effects. Intraclass correlation coefficients (ICCs) were computed to identify how much of the total variance in health care costs can be attributed to the care groups and GPs (Snijders and Bosker 2012). The ICCs at GP level were calculated by dividing the variation at GP level by the total variation. The ICCs at care group level were calculated by dividing the variation at care group level by the total variation (Hox 2002). Patients with zero costs were excluded from the analyses. All 104,544 patients had more than zero total treatment health care costs, but only 9,949 had made diabetes-specific specialist care costs. Appendix S3 provides more information on the estimation procedures and alternative model specifications.

To answer the second research question regarding the GLMM approach versus the simple mean aggregation method, we compared the two methods of ranking. The first method consists of plotting the unadjusted means of health care costs per care group and their corresponding CIs. In the current study, we report on a 95 percent confidence level, that is, the sample estimate of the mean $\pm 1.96 \frac{\text{St. Dev.}}{\sqrt{n}}$. Note, however, that this assumes the sample distribution to be symmetric and without outliers (for small sample sizes) or moderately skewed for larger sample sizes; in summary, although better than only providing a point estimate of the mean, this method does not perform well for smaller sample sizes per care group or a highly skewed cost distribution. The second method consists of plotting the predicted estimates of care group effects, also known as “posterior means,” “predicted residual errors,” or

shortly “residuals” (Hox 2002) and their corresponding CIs of the three-level random-effects generalized linear model which allows for nonnormal error distributions. Finally, we created a scatterplot of both ranking methods to demonstrate the differences in rankings caused by the methods used. Reliability per care group was calculated by first calculating the GP-reliability ($=\lambda$ per GP) and using the sum of all GP λ ’s per care group to calculate the care group λ (Raudenbush and Bryk 2002). We provide more detailed information on calculating the reliability of aggregated variables in Appendix S3. Moreover, to study the strength of the linear association between both measurements, we computed the Pearson correlation coefficient in Excel. All analyses were performed using SAS 9.3.

RESULTS

Characteristics of Study Population

Complete insurance claim data of 104,544 eligible diabetes type 2 patients from 2,067 GPs participating in 50 care groups for the year 2009 were used (see Appendix S2A). The average number of diabetes patients per care group was more than 2,000 but varied widely from 15 to more than 8,000 patients. Half of the patients were female and the average age was about 67 years.

Treatment health care costs per diabetes patient were on average about €4,500 (Appendix S2A). General practice costs constituted only a limited share of that figure, about €500. The largest cost items were for specialist care (averaging about €2,000 per patient), with only €63 of specialist care costs being diabetes-specific. Patients that had consumed diabetes-specific specialist care had on average higher total treatment health care costs (about €7,000; Appendix S2B).

Effects of Patient Characteristics

Table 1 presents the model estimates of the 2009 total treatment health care costs (Model 1 and 2) and diabetes-specific specialist care (Model 3 and 4) of diabetes patients in which patient characteristics and clustering of data in higher levels were taken into account. Presented results refer to the expected change in the natural logarithm of Y. Table 1, Model 1 shows that the average patient across GPs and care groups made costs of €4,046 ($=\exp(7.8788) \cdot \exp(0.8534/2)$). Model 3 shows the diabetes-specific specialist care costs for the average patient were €587 ($=\exp(6.1653) \cdot \exp(0.4202/2)$).

Table 1: Effects of Patient, GP, and Care Group Clustering on Total Curative Health Care Costs and Diabetes-Specific Specialist Care Costs of Adult Type 2 Diabetes Patients Enrolled in Disease Management Programs of Care Groups in 2009

	Total Curative Health Care Costs (N = 140,544)			Diabetes-Specific Specialist Care Costs (N = 9949)		
	Model 1	SE	Model 2	SE	Model 3	SE
Intercept	7.8788	0.0008	7.6817	0.0102	6.1653	0.02669
Explanatory variables						
Age (centered at 67.1 year)			0.009596	0.0002		0.0013
Sex (ref. = female)			0.02343	0.0054		0.0260
Number of comorbid conditions (1, 2, 3 or more, ref. = 0)			-0.5288	0.0061		0.0549
			-0.9564	0.0106		0.0811
			-1.5218	0.0745		-0.0607
Variance per level						
Patient	0.8534	0.0038	0.7320	0.0032	0.4202	0.0065
GP	0.0087	0.0008	0.0069	0.0007	70.0041	0.0027
Care group	0.0031	0.0010	0.0032	0.0009	0.0278	0.0064
Intraclass correlation (%)						
GP	1.18		0.44		0.98	
Care group	0.00		0.38		6.62	
-2 log likelihood	280,979.2		264,900.6		19,807.4	

Notes: GP = general practitioner; SE = standard error. p -value $\leq .05$; p -value $\leq .01$; p -value $\leq .001$. Model 1: Null model for total curative health care costs; Model 2: Model for total curative health care costs with explanatory variables; Model 1 and 2: Number patients = 104,544; Number of GP's = 2,067; Number care groups = 50; Average number patients per care group = 2,091; Total curative health care costs were defined as the sum of general practice costs, costs of specialist care, and other curative costs (i.e., costs for pharmaceutical care, medical aids, physiotherapy, exercise therapy, speech therapy, occupational therapy, dietetics, patient transport, and mental health care) per patient; Model 3: Null model for diabetes-specific specialist care costs; Model 4: Model for diabetes-specific specialist care costs with explanatory variables. Model 3 and 4: Number patients = 9,949; Number of GP's = 1,828; Number care groups = 47; Average number patients per care group = 212; Diabetes-specific specialist care costs were defined as the sum hospital costs per patient with diagnosis diabetes of outpatient, day-patient, and inpatient care.

The age and sex of patients and the number of comorbid conditions were significant confounder variables in Model 2. Patient characteristics were associated with diabetes-specific specialist care costs (except “sex” and one category of comorbidity as both are not significant in Model 4). Because of the confounder variables, the total treatment health care costs and the diabetes-specific specialist care costs of a “typical” patient across GPs and care groups decreased with, respectively, 22 and 3.5 percent.

Effects of Clustering of the Data

In the model with the first indicator the variance on the GP level was small and on the care group level negligible compared to the patient-level variance (Table 1). The influence of the care group was stronger on the more care group-specific dependent variable, diabetes-specific specialist care costs (ICC >6 percent), while the GP-level variation was rather small (1 percent). As including patient characteristics showed no changes in the variances (Table 1, Model 3 and 4), the 6 percent care group-level variance is not a composition effect of these characteristics.

Ranking Care Groups by Different Methods

In Figure 1, care groups were ranked after taking the simple mean of the total treatment health care costs per patient per care group. The simple means of the 2009 total treatment health care costs varied between care groups from about €3,100 to €6,500. These are the “huge cost differences between care groups” that previous studies have found. Based on these health care cost differences between care groups, one could indeed argue that large efficiency gains are potentially achievable. However, the CIs of these simple means were so high that no significant variation between care groups could be observed.

Figure 2 ranks care groups by plotting their residuals, which account for the clustered structure of the data and several patient characteristics (derived from data of Model 2 in Table 1, GLMM approach). Figure 2 shows that some care groups significantly differed from the grand mean across all care groups (i.e., their CIs do not cross the x -axis), although the care group variance in Table 1 Model 2 showed to be very small. The typical patient of the “best performing” care group had €2,884 ($=\exp(7.6817 + (-0.08071)) \cdot \exp(0.732/2)$) total treatment health care costs, while the “worst performing” care groups had €3,511 ($=\exp(7.6817 + (0.116))$)

Figure 1: Care Groups Ranked* by the Simple Mean Aggregation of Curative Health Care Costs (€) of Their Diabetes Patients in 2009, Including Confidence Intervals ($N = 52$)

Notes: *Care group numbers on x-axis are derived from the electronic identification number of care groups; care groups are randomly pseudonymised.

* $\exp(0.732/2)$). This rather small difference of €400 would decrease even further if other patient- and GP-level characteristics could be included to account for more composition effects.

To answer research question 2, care groups were first ranked by using the simple mean of diabetes-specific specialist care costs of their patients that had consumed this kind of care (Appendix S4). When applying the GLMM approach with the dependent variable diabetes-specific specialist care costs (Table 1, Model 4), significant variation between care groups was observed (Figure 3). More than a third of the care groups diverged from the grand mean of all care groups (their CIs do not cross the x-axis). Nine care groups had significantly higher than average diabetes-specific specialist care costs, while for 11 care groups these costs were significantly lower compared to the grand mean of all care groups. The care group with the random number 1 is an extreme outlier; while the typical patient across care groups and GPs had dia-

Figure 2: Care Groups Ranked by the Residuals (Including Confidence Intervals) of a GLMM with a Log-Normal Distribution and Identity Link Function of Curative Health Care Costs of Diabetes Patients in 2009, Adjusted for Patient Characteristics and Clustering of Patients in GPs and GPs in Care Groups ($N = 50$)

betes-specific specialist care costs of €566 ($=\exp(6.1301)*\exp(0.4186/2)$) and the costs of the typical patients for the “worst” care group were €997 ($=\exp(6.1301 + (0.565))*\exp(0.4186/2)$).

To demonstrate the difference in ranking order of care groups between both methods, we plotted the care groups ranks based on the second cost indicator. We plotted the ranking of the care groups for both methods in a scatter-plot (Figure 4), with on the X-axis the care groups ranking number based on the simple mean aggregation (derived from Appendix S4), and on the Y-axis the ranking number of the same care group based on the GLMM approach (derived from Figure 3). Rank 1 refers to the care group with the lowest cost residual (the best performer) and rank 47 refers to the care group with the highest cost residual (the worst performer). If the chosen method would have a limited or no effect on the care groups’ ranking order, all points would be close to or on the 45° line.

Figure 4 shows that some points (circle symbols) are indeed close to the 45° line, showing less than 5 rank order differences. The Pearson correlation

Figure 3: Care Groups Ranked by the Residuals (Including Confidence Intervals) of a GLMM with a Log-Normal Distribution and Identity Link Function of Diabetes-Specific Specialist Care Costs of Diabetes Patients in 2009, Adjusted for Patient Characteristics and Clustering of Patients in GPs and GPs in Care Groups ($N = 47$)

Notes: Patients with zero costs were excluded. $N = 9,949$ patients. * Confidence intervals that cross the x -axis cannot be distinguished significantly from the mean care group costs.

between the two methods was fairly high (0.79), too. However, 27 care groups showed a rank difference in methods of five positions or more (triangle and diamond symbols). A fair comparison of methods only compares rankings of care groups with reliably predicted residual point estimates (filled symbols). Of all care groups with reliable information (lambda higher than 0.6), 14 differed between five and nine positions (filled triangle symbol), and 5 differed 10 or more positions (filled diamond symbol). In summary, the statistical method used strongly affects the ranking pattern of care groups.

Differences in rank position depending on the method of aggregation lead to differences in the upper and lower top-10 categories. Figure 4 shows that four (two reliable) of the top-10 of the simple ranking have lower positions in the GLMM ranking. Six (three reliable) care groups are not part of the lower top-10 in the GLMM method, while they are in the simple mean aggregation.

Figure 4: Scatterplot of Ranking of Care Groups Based on the Diabetes-Specific Specialist Care Costs of Diabetes Patients with More Than Zero Costs, Calculated by the Simple Mean Aggregation (*X*-axis) and a GLMM with a Log-Normal Distribution and Identity Link Function (*y*-axis)

Notes: λ = Lambda, reliability of the care group level cost estimate. diff = difference in position between two methods of ranking.

DISCUSSION

The aim of this study was twofold. First, we aimed to investigate the association between the newly introduced care groups and health care costs of diabetes type 2 patients in general. We conclude that care groups—next to GPs—appear to have a significant but small context effect on total health care costs per diabetes patient. This might be because overall cost variables are likely to be biased by several confounders (like nondiabetes-specific health care costs). Therefore, these overall cost variables may not be suitable for identifying care

group variation and performance, although a reduction in the total costs is a desirable goal to achieve. As care groups have a negligible effect on health care spending of their patients, it is questionable whether care groups in their current form will realize the savings as projected by the Ministry of Health, Welfare and Sports (CPB Netherlands Bureau for Economic Policy Analysis 2009), or even achieve any slowdown in the cost growth. Furthermore, this study showed a care group context effect on diabetes-specific specialist care costs per patient. The patients' characteristics (the composition for which we were able to control) does not explain the context effect of care groups. Thus, this study showed that care groups do vary with regard to diabetes-specific specialist care costs. Our results suggest that when cost variation is studied, it is desirable to choose a cost indicator that has the potential to be directly influenced by care groups.

The second aim of this study was to determine whether ranking of care groups using the GLMM approach differs from ranking care groups using a simple mean aggregation. The difference between the simple mean aggregation and the GLMM approach seems small because of the high correlation, and one might prefer the easier applicable simple mean aggregation method. However, we also showed that this will not be without consequences for specific care groups. If cut-off points are going to be used to penalize or reward the performance of care groups, the exact ranking position will be very important for care groups, and by extension the method of aggregation. Hence, our study implies that careful consideration is required on, first, what type of health care costs to include and, second, what kind of statistical method to use in studying the performance of care groups. We recommend the GLMM approach or similar approaches for a more precise ranking to investigate the performance of care groups and diabetes-specific or diabetes-related specialist care costs as outcome variables.

Our findings are not in agreement with other Dutch publications (Achmea 2011; Van Schooten et al. 2012) on this topic, which stated that financial performance of "total costs" largely differed between care groups. However, we do conclude that some significant differences between care groups exist, when an appropriate indicator and a more sophisticated aggregation technique is used.

Importantly, this study did not aim to investigate *how* care groups affect health care costs, for example, by improving quality of care or facilitating a more efficient care delivery process. Although these are the expected mechanisms, we discuss several other possibilities. For example, the Dutch pricing system for specialist care differs between hospitals; physicians are either self-

employed professionals and mainly paid via fee-for-services principles, or they receive a fixed salary from the hospital. Salaried physicians produce less volume than physicians reimbursed via fee-for-service (Douven, Mocking, and Mosca 2012). Next to the payment system of hospitals, the traveling distance between the patient and the hospital is also of importance for secondary health care use and costs. The closer patients live to a hospital, the higher their preference for that hospital and the more likely it is that their GPs will refer them to that hospital (Birk and Henriksen 2012). Consequently, care groups situated in the catchment area of an “expensive” hospital will have higher (diabetes-specific) specialist care costs as compared to care groups situated nearby a “cheaper” hospital or care groups situated in a more rural area. Another, more technical, explanation might be that hospitals differ in their coding practices. For instance, classifying health care claims procedures as “diabetes-specific” or “diabetes-related” like cardiovascular complications can be arbitrary and associated with other incentives and therefore might differ across hospitals (Hasaart 2011). Hence, these hospital-related differences regarding (up-)coding procedures might have contributed to the care group variation without reflecting care group performance. Another reason for differences in diabetes-specific specialist care might be due to patient characteristics unknown to us that influence their decisions regarding health care utilization.

Our study has some limitations. Health claims data lack several patient characteristics that may affect the health care expenditures, like socio-economic status (Epstein, Stern, and Weissman 1990) and health-related behavior (Sturm 2002). Consequently, we were not able to adjust for these compositional factors. Therefore, care groups situated in areas with high rates of people with a low socio-economic status might have higher per-patient costs that are incorrectly assigned to the care group performance. As most care groups cover a geographic area, the care group in which a patient is enrolled depends on where patients live, rather than on patient preferences. This implies that selection effects due to, for example, morbidity-related factors, on top of socio-economic and demographic variation, are very unlikely. Care groups variance may have been overestimated by not accounting for this and other covariates with which context can be more adequately separated from composition (Bingenheimer and Raudenbush 2004). Moreover, we confined our study to care groups that started up before 2009, and thereby only included the more frontrunner care groups. These frontrunners might perform differently than the later established care groups. Future research will be able to investigate the performance of all 100 Dutch care groups.

Some implications for policy makers and future research arise from our results. First, at this early stage of care groups and BPs, it is of utmost importance to develop scientifically sound methods for identifying care group variation for all stakeholders among which insurers, the Dutch Health Care Inspectorate, Dutch citizens, and care groups themselves. This study showed that the ranking of care groups depends on both the methods and the indicators used. By using inappropriate methods and indicators, the wrong care groups may be targeted to improve their performance and may experience negative consequences of these rankings during their negotiations with insurers.

Moreover, our results are also important for other countries in which provider-led entities, comparable to care groups, will be increasingly benchmarked in the coming years. For example, England's NHS has made a first attempt to benchmark the Clinical Commissioning Group (CCG) with an "information pack." This information pack allows CCGs to be benchmarked on several indicators with the purpose to inform CCGs about their current position compared to other CCGs (NHS 2012). Also, the US government established a national Accountable Care Organization benchmark for 33 different quality measures (Centers for Medicare & Medicaid Services, 2014). Moreover, our findings are likely to be applicable not only for care groups but also for other health care providers.

Finally, to improve the ranking process of care groups and quality of care in general, further research should include—next to more information on hospitals in the adherence area—the content of the DMPs, patient characteristics such as socio-economic status, and outcome indicators on quality of care. More reliable conclusions about the cost-effectiveness of care delivered by care groups are only possible by combining health claims data, quality data, and information on the content of the DMP. Future research should also pay special attention to causes of variation between care groups in terms of quality, costs, and content of the services they deliver.

CONCLUSION

Our study showed that the care groups' performance differs to a limited extent with regard to total treatment health care costs per patient and to a larger extent with regard to diabetes-specific specialist care costs per diabetes patient. Next to that, benchmarking care groups based on taking the simple mean of health care expenditures across all patients probably have unfair consequences for some care groups.

ACKNOWLEDGMENTS

Joint Acknowledgment/Disclosure Statement: The authors thank Vektis for providing access to the health claims data and for assisting with their comprehensive knowledge regarding health claims data. The authors also thank Albert Wong and Maarten Schipper for their suggestions for data analysis. During preparation of this manuscript Jeroen N. Struijs held a Harkness Fellowship of the Commonwealth Fund. The views expressed are those of the authors and do not reflect those of the Commonwealth Fund or their staff.

Disclosures: None.

Disclaimers: None.

REFERENCES

- Achmea. 2011. *Kwantitatieve inzichten van 4 jaar ervaring met integrale bekostiging diabeteszorg en landelijke ervaringen met diabetesketenzorg en kwalitatieve resultaten [Quantitative Insights of 4 Years Experience with Integrated Diabetes Care Funding and National Experiences with Diabetes Integrated Care and Qualitative Results]*. Amersfoort, The Netherlands: Achmea, Department Health Care and Health.
- Adviesgroep Ketenzorg. 2012. *Transparantie ketenzorg Diabetes Mellitus. Rapportage zorggroepen 2010. Op weg naar genuanceerde rapportage van zorg [Transparency of Integrated Care for Diabetes Mellitus. Reporting Care Groups in 2010. Towards Nuanced Reporting of Care]*. Utrecht, the Netherlands: Adviesgroep Ketenzorg.
- Berenson, R. A., P. J. Pronovost, and H. M. Krumholz. 2013. "Achieving the Potential of Health Care Performance Measures." In *Timely Analysis of Immediate Health Policy*, pp 1–31. Robert Wood Johnson Foundation Urban Institute [accessed on July 3, 2015, 2013]. Available at <http://www.rwjf.org/content/dam/farm/reports/reports/2013/rwjf406195>
- Bingenheimer, J. B., and S. W. Raudenbush. 2004. "Statistical and Substantive Inferences in Public Health: Issues in the Application of Multilevel Models." *Annual Review of Public Health* 25: 53–77.
- Birk, H. O., and L. O. Henriksen. 2012. "Which Factors Decided General Practitioners' Choice of Hospital on Behalf of Their Patients in an Area with Free Choice of Public Hospital? A Questionnaire Study." *BMC Health Services Research* 12: 126.
- Centers for Medicare & Medicaid Services. 2014. "Medicare Shared Savings Program Quality Measure Benchmarks for the 2014 Reporting Year" [accessed on July 3, 2015]. Available at <http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/sharedsavingsprogram/Downloads/MSSP-QM-Benchmarks.pdf>
- Christiansen, C. L., and C. N. Morris. 1997. "Improving the Statistical Approach to Health Care Provider Profiling." *Annals of Internal Medicine* 127 (8 Pt 2): 764–8.
- Clark, T. S., and D. A. Linzer. 2014. "Should I Use Fixed or Random Effects?" *Political Science Research and Methods* 3 (2): CJO2014. doi:10.1017/psrm.2014.32.

- CPB Netherlands Bureau for Economic Policy Analysis. 2009. *Houdbaarheidsmaatregelen curatieve zorg [Rules of Stability in Curative Care]*. The Hague: CPB Netherlands Bureau for Economic Policy Analysis.
- De Bakker, D. H., J. N. Struijs, C. B. Baan, J. Raams, J. E. de Wildt, H. J. Vrijhoef, and F. T. Schut. 2012. "Early Results from Adoption of Bundled Payment for Diabetes Care in the Netherlands Show Improvement in Care Coordination." *Health Affairs (Millwood)* 31 (2): 426–33.
- De Bo, A. 2011. "Vektis 'Informatiecentrum voor de zorg' [Vektis 'Information Center for Health Care']." *Tijdschrift voor sociale gezondheidszorg* 89 (7): 358–9.
- De Bruin, S. R., C. A. Baan, and J. A. N. Struijs. 2011. "Pay-for-Performance in Disease Management: A Systematic Review of the Literature." *BMC Health Services Research* 11: 272.
- De Jong-van Til, J., L. Lemmens, C. Baan, and J. Struijs. 2012. *De organisatie van zorggroepen anno 2011: Huidige stand van zaken en de ontwikkelingen in de afgelopen jaren [The Organization of Care Groups in 2011: Current State of Affairs and Development in Recent Years]*. Bilthoven: National Institute for Health and the Environment.
- Dieleman, J. L., and T. Templin. 2014. "Random-Effects, Fixed-Effects and the within-between Specification for Clustered Data in Observational Health Studies: A Simulation Study." *PLoS ONE* 9 (10): e110257.
- Doran, T., E. Kontopantelis, J. M. Valderas, S. Campbell, M. Roland, C. Salisbury, and D. Reeves. 2011. "Effect of Financial Incentives on Incentivised and Non-Incentivised Clinical Activities: Longitudinal Analysis of Data from the UK—Quality and Outcomes Framework." *British Medical Journal* 343 (7814): 342:d3590. doi:10.1136/bmj.d3590.
- Douven, R., R. Mocking, and I. Mosca. 2012. *The Effect of Physician Fees and Density Differences on Regional Variation in Hospital Treatments*. CPB Discussion Paper. The Hague: CPB Netherlands Bureau for Economic Policy Analysis [accessed on July 3, 2015]. Available at http://www.nza.nl/104107/230942/Research_Paper_The_Effect_of_Physician_Fees_and_Density_Differences.pdf
- Duan, N. 1983. "Smearing Estimate: A Nonparametric Retransformation Method." *Journal of the American Statistical Association* 78 (383): 605–10.
- Duncan, C., K. Jones, and G. Moon. 1998. "Context, Composition and Heterogeneity: Using Multilevel Models in Health Research." *Social Sciences & Medicine* 46 (1): 97–117.
- Epstein, A. M., R. S. Stern, and J. S. Weissman. 1990. "Do the Poor Cost More? A Multihospital Study of Patients' Socioeconomic Status and Use of Hospital Resources." *New England Journal of Medicine* 322 (16): 1122–8.
- Gregori, D., M. Petrinco, S. Bo, A. Desideri, F. Merletti, and E. Pagano. 2011. "Regression Models for Analyzing Costs and Their Determinants in Health Care: An Introductory Review." *International Journal for Quality in Health Care* 23 (3): 331–41.
- Hasaart, F. 2011. *Incentives in the Diagnosis Treatment Combination Payment System for Specialist Medical Care*. Dissertation. Maastricht: University of Maastricht.
- Hox, J. 2002. *Multilevel Analysis: Techniques and Applications*. Mahwah, NJ: Lawrence Erlbaum Associates.

- Iezzoni, L. I., A. S. Ash, M. Schwartz, J. Daley, J. S. Hughes, and Y. D. Mackiernan. 1996. "Judging Hospitals by Severity-Adjusted Mortality Rates: The Influence of the Severity-Adjustment Method." *American Journal of Public Health* 86 (10): 1379–87.
- Kilian, R., H. Matschinger, W. Loeffler, C. Roick, and M. C. Angermeyer. 2002. "A Comparison of Methods to Handle Skew Distributed Cost Variables in the Analysis of the Resource Consumption in Schizophrenia Treatment." *Journal of Mental Health Policy and Economics* 5 (1): 21–31.
- Landelijke Organisatie voor Keten zorg. 2013. *Transparante ketenzorg Diabetes Mellitus, COPD en VRM rapportage zorggroepen over 2012 [Report on Transparency of Care Groups of Integrated Care for Diabetes Mellitus, COPD and VRM 2012]*. Utrecht, the Netherlands: Landelijke Organisatie voor Keten zorg.
- Litière, S., A. Alonso, and G. Molenberghs. 2007. "Type I and Type II Error under Random-Effects Misspecification in Generalized Linear Mixed Models." *Biometrics* 63: 1038–44.
- Malmström, M., J. Sundquist, and S. E. Johansson. 1999. "Neighborhood Environment and Self-Reported Health Status: A Multilevel Analysis." *American Journal of Public Health* 89 (8): 1181–6.
- Manning, W. G., and J. Mullahy. 2001. "Estimating Log Models: To Transform or Not to Transform?" *Journal of Health Economics* 20 (4): 461–94.
- Mohnen, S. M., P. P. Groenewegen, B. Völker, and H. Flap. 2011. "Neighborhood Social Capital and Individual Health." *Social Science & Medicine* 72 (5): 660–7.
- NHS. 2012. "Outcomes Benchmarking Support Packs: CCG Level." National Health Service [accessed on July 3, 2015]. Available at <http://www.england.nhs.uk/wp-content/uploads/2012/12/ccg-pack-02p.pdf>
- Raudenbush, S. W., and A. S. Bryk. 1986. "A Hierarchical Model for Studying School Effects." *Sociology of Education* 59: 1–17.
- . 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. Newbury Park, CA: Sage Publications.
- Raudenbush, S. W., and R. J. Sampson. 1999. "Econometrics: Toward a Science of Assessing Ecological Settings, with Application to the Systematic Social Observation of Neighborhoods." *Sociological Methodology* 29 (1): 1–41.
- Salem-Schatz, S., G. Moore, M. Rucker, and S. D. Pearson. 1994. "The Case for Case-Mix Adjustment in Practice Profiling: When Good Apples Look Bad." *Journal of the American Medical Association* 272 (11): 871–4.
- Snijders, T. A. B., and R. J. Bosker. 2012. *Multilevel Analysis. An Introduction to Basic and Advanced Multilevel Modeling*, 2nd Edition. London: SAGE Publications.
- Song, Z., D. G. Safran, B. E. Landon, Y. He, R. P. Ellis, R. E. Mechanic, M. P. Day, and M. E. Chernew. 2011. "Health Care Spending and Quality in Year 1 of the Alternative Quality Contract." *New England Journal of Medicine* 365 (10): 909–18.
- Struijs, J. N., and C. A. Baan. 2011. "Integrating Care through Bundled Payments—Lessons from the Netherlands." *The New England Journal of Medicine* 364 (11): 990–1.
- Sturm, R. 2002. "The Effects of Obesity, Smoking, and Drinking on Medical Problems and Costs." *Health Affairs* 21 (2): 245–53.

- Tsiachristas, A., C. Dijkers, M. R. S. Boland, and M. P. M. H. Rutten-van Mölken. 2013. "Exploring Payment Schemes Used to Promote Integrated Chronic Care in Europe." *Health Policy* 13 (3): 296–304.
- Van Herck, P., D. De Smedt, L. Annemans, R. Remmen, M. B. Rosenthal, and W. Sermeus. 2010. "Systematic Review: Effects, Design Choices, and Context of Pay-for-Performance in Health Care." *BMC Health Services Research* 10: 247. doi:10.1186/472-6963-10-247.
- Van Schooten, G., M. Van Galen, P. Sterkenburg, N. Beersen, and M. Berg. 2012. *Kwaliteit en kosten van de geleverde zorg rond COPD en astma in 2010. Analyse van de waarde van COPD- en astmazorg op basis van declaratiegegevens [Quality and Cost of Care Delivered for COPD and Asthma in 2010. Analysis of the Value of COPD and Asthma Care Based on Health Claims Data]*. Zeist: Zorgverzekeraars Nederland.
- Wennberg, J. E. 2011. "Time to Tackle Unwarranted Variations in Practice." *British Medical Journal* 342: 687–690.
- Zorgverzekeraars Nederland. 2011. *Rapportage indicatoren indicatiestelling (praktijkvariatie) [Reporting Indicators Assessment (Practice Variation)]*. Zeist: Zorgverzekeraars Nederland.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.

Appendix S1: Reasons for Exclusion and Selection of Patients.

Appendix S2A: Characteristics of the Study Population ($N = 104,544$).

Appendix S2B: Characteristics of the Study Population ($N = 9,960$).

Appendix S3: Additional Information on Statistical Analyses.

Appendix S4: Care Groups Ranked by the Simple Mean Aggregation of Diabetes-Specific Specialist Care Costs (€) of Their Diabetes Patients with Diabetes-Specific Specialist Care Costs ($n = 9,949$) in 2009 (Including Confidence Intervals) ($N = 47$).