Performance differences between the episode-based DBC and diagnosis-related DRG case mix systems

Yvonne Krabbe-Alkemade en Tom Groot

SUMMARY This paper explores the question how much detail a cost system needs to have in order to provide reliable cost information at a reasonable price. In general, fine-grained cost systems with a lot of detail (in product definition, in cost drivers and in cost pools) are expected to provide more reliable cost information than coarse-grained cost systems with less detail. This paper takes as an example the DBC cost system that has been developed for the Dutch hospital sector. The fine-grained DBC system with over 40,000 health care products appears to outperform lower-grained DRG systems with "only" 15,000 and 6,000 health care products on cost homogeneity and predictive validity. It does so however at the cost of a high number of products with measurement and specification errors, caused by a large number of outliers and by a low number of observations in product groups. The cost-effectiveness of the DBC system is not very high: only 3% of all DBC-codes explains 80% of total costs, whereas the lower-grained DRG system uses 14% of the codes to explain 80% of total costs. Combined with the high administration cost of the DBC-system, it was from an economic perspective, a sensible idea to replace the fine-grained DBC-system by the coarse-grained DOT system.

PRACTICAL RELEVANCE More detailed cost systems are not necessarily "better" cost systems in terms of cost-benefit and accuracy of cost information.

1 Introduction

In the last forty years, most countries around the world have put a lot of time and effort into developing systems that make visible what takes place in the hospital so that hospital production can be measured and evaluated in a systematic way (Busse et al., 2011). Most countries have by now adapted a case mix system in which patients are classified in different categories which are homogeneous in medical terms and meaningful in economic terms. The most well-known case mix system is the HCFA (Health Care Financing Administration) patient classification system, or Diagnosis Related Groups system (DRG), which was introduced by Fetter in 1983 to facilitate the Medicare prospective payment system. Since its inception, many countries have adopted the DRG system or implemented an adapted version of it to suit local requirements (Fetter et al., 1991). Although there are many studies about the effects of DRG systems on health care costs and care provision, the impact of different design characteristics of case mix systems on decision making and on the cost of information has not yet been extensively studied (Quentin et al., 2011).

The first version of the Dutch case mix system was not developed as an adaptation of the Yale DRG system, but was designed as a completely new system. This system has become known as the *Diagnosis Treatment Combination* system (in Dutch "Diagnose Behandeling Combinatie": DBC) and contained more than 40,000 different care products. This allows for a high level of detail which makes DBC a fine-grained cost information system. In fact, the DBC system is far more fine-grained than any other DRG system currently in operation. Most DBC systems contain between 600 and 2,300 care products (Kobel et al., 2011). The development of a different and much more fine-grained patient classification system in the Netherlands was caused by a combination of design choices and the decision-making processes used in the development of the DBC system. The DBC system is currently no longer in operation: it has served between 2005 and 2011 and was replaced in 2012 by the DOT system. The abbreviation DOT stands for "DBCs becoming more transparant" (in Dutch: "DBC's op weg naar transparantie").

The DBC and DRG systems differ in some important aspects (Steinbusch et al., 2007; Westerdijk et al., 2012; Schreyogg et al., 2006; Busse et al., 2011). A DRG is based on an inpatient episode summary. Some systems also use day care episodes, e.g., the German DRG system. The DBC system contains clinical process summaries by episode of care including outpatient visits, clinical episodes, day-care and after-care. It also includes information about the diagnosis, type of care delivered (initial or follow-up care) and type of care demanded (by general physician or specialist). In the DBC system, one referral may lead to more episodes of care or even to new DBCs, in case of co-morbidity. The DRG systems are mostly restricted to one classification for each clinical episode. Medical administrators do DRG coding after a patient's dismissal. DBC coding is done by or under supervision of clinicians during the health care process. Most DRG systems include the fee for medical specialists, while this fee is separately registered in the DBC system. Finally, the DRG system is linked to an inpatient admission system, leading to a single invoice. Under the DBC system, an episode of care can be described by more than one DBC health product, leading to several DBC invoices for a specific care episode (Westerdijk et al., 2012).

The DBC system appeared to be too fine-grained for use as a negotiation tool between care providers and care insurers. It did not lead to a meaningful grouping of health care products and it led to excessive administrative costs (DBC-Onderhoud, 2007a, 2007b; NZa, 2011). The current DOT system consists of 4,400 care products and their identifications are based on the ICD-10 classification of diagnoses. The DOT system leads to definitions of health care products that are much more similar to patient classifications in other DRG systems. This also facilitates international comparisons, coordination and charging of patients across borders. This development shows that the Dutch DBC system and international developments in DRG systems converge (Busse, et al., 2011). The new Dutch DOT system is still more fine-grained than most other DRG systems, but the extremely fine-grained DBC system no longer exists.

The abolishment of the fine-grained DBC system seems mainly motivated by the desire to simplify the case mix system, in order to make it more useful for contracting and internal management purposes, to lower administration costs, and to make DBC-information more internationally comparable. What remains unclear, however, is to what degree the abolishment of the DBC system has led to less-reliable resource utilization information. One could expect a more fine-grained case mix system to produce more accurate cost information than would a less fine-grained case mix system. Cost information accuracy is higher when unit cost information represents more fairly resources consumed. On the other hand, fine-grained systems that fail to define economically homogeneous patient groups may not produce equally reliable cost information when compared to less fine-grained systems using more cost-homogeneous patient groups.

This paper evaluates the quality of the patient classification system among different granularity levels. We use the original DBC product structure as an example of the fine-grained product classification. We contrast this system with information generated by the same system, using only the diagnoses information and thus leaving out the episode-of-care information. By doing so, we reduce the granularity level from 44,000 to 2,300 products, which is an granularity level that is comparable to most of the existing DRG systems elsewhere. We also use two alternative aggregation methods: one based on a combination of diagnoses and treatments (14,000 products), and another based on a combination of diagnoses and type of care (6,400 products). The granularity level of these product aggregations lies in between those of DBC and diagnose-based aggregation.

Following common-sense reasoning, more fine-grained cost systems can generally be expected to portray product costs more accurately than coarse-grained cost systems. However, errors of measurement, aggregation and specification that may occur in the process of refining the cost system could eventually lead to less accurate, instead of more accurate, product cost information. Since we do not know the accurate product costs, we do not have the appropriate benchmark to assess the accuracy of the different case mix systems. Instead, we evaluate the performance of the case mix systems with different granularity levels using three important cost system characteristics: cost-effectiveness, within-product homogeneity and predictive validity. Our conclusion is that the fine-grained Dutch DBC system is not very cost-effective, but it outperforms other systems on within-product homogeneity and predictive validity. However, the DBC system in its original form (in 2007) also had the propensity to compound the effects of measurement and aggregation errors. This introduces the possibility that the DBC system does not produce more accurate case mix cost information than more coarse-grained information systems, such as those based on diagnostic information.

This paper is organized as follows. Section 2 describes the relevant theoretical framework as a basis for our research question. In Section 3, the research methodology is described. Section 4 gives a description of the data. Section 5 provides the analysis and results. The last section contains a discussion of the results and describes several conclusions, including suggestions for further research.

2 Granularity and quality of product costing systems

Cost information can be used for a variety of purposes, like documenting the cost of a particular illness or treatment, assigning financial responsibilities to decision-makers in hospitals, benchmarking costs with other suppliers, estimating the costs of health care innovations, and determining prices in prospective payment systems. How fine-grained a patient classification needs to be depends on the balance between information costs and decision benefits (Jackson, 2000). In this study, we compare different granularity levels of the Dutch payment system using three evaluation criteria: (1) cost effectiveness, (2) within-group homogeneity of case mix classes, and (3) predictive validity of the case mix system.

Standard costing theory posits that the use of crude proxies of resource consumption (e.g. 'length-of-stay') for costing purposes, under conditions of cost heterogeneity and different resource usage patterns, may lead to distorted cost information (Cooper & Kaplan, 1988, 1992). Product costing systems may more accurately capture resource consumption patterns when they are based on different resource and activity cost pools, and when resource consumption is traced to products by cost drivers that more adequately represent resource consumption patterns. Choosing among alternative systems of health care product costing can be considered an exercise in minimizing product-costing errors. Product cost estimate errors come in three categories: measurement errors, aggregation errors and specification errors (Datar & Gupta, 1994; Labro & Vanhoucke, 2007; Gupta, 1993).
Measurement errors originate when costs and related variables, like costs allocation drivers, are not supported by well-defined measurement techniques and measurement guidelines, including specifications of cost items. The use of length of stay as a proxy for costs may cause measurement error when treatment costs are not uniformly distributed over hospital days. Expert opinions used for costing purposes may also lead to significant measurement errors. Systems such as the DBC system, which uses cost information that is derived from the aggregation of care activities, are supposed to be less prone to measurement error than systems using proxies for or subjective estimations of total costs. This, however, assumes that DBC categories are easily recognized by caregivers and that activities can be measured with sufficient precision.
Aggregation errors occur when heterogeneous costs or resource cost pools are accumulated in a single activity cost pool or when a single cost allocation rate is applied over heterogeneous activities. The use of a larger set of different cost pools and allocation rates for the allocation of hospital costs over health care products reduces the risk of aggregation errors.
Specification errors arise when cost driver units do not reliably reflect the demands placed on resources by individual products. These errors may occur in two instances: by mis-specifying resources to activities (resource drivers) and activities to products (activity drivers). A specification error commonly occurs when costs do not vary directly with volume, e.g., setup-costs or batch-related costs.

The accuracy of health care product costs is a function of complex interactions among the three types of errors. Case-based evidence shows that a more fine-grained cost system may not always lead to more accurate product cost figures (Datar & Gupta, 1994; Gupta, 1993). Under certain conditions, aggregation, specification and measurement errors may (partially) offset each other, which may lower the rate of error in more aggregate costing systems. When errors in more aggregate costing systems (partially) cancel each other out, they may even increase total error in product costing when a more refined cost model is used. Christensen and Demski (1995) make a similar point. They note that the use of multiple cost pools, aimed at reducing aggregation errors, may eventually lead to less accurate product costs. That is, the use of more cost pools may lead to higher measurement errors, offsetting the error reduction from using a more refined set of cost pools.
Simulations of two-stage cost allocation models have shown that in general, incremental refinements of the allocation system do lead to overall improvement of total cost information accuracy. However, some offsetting mechanisms also appear to exist: measurement errors in resource cost pools have greater potential to be offset when activity cost pools are more aggregated. Cost system refinements lead to more accurate total product costs when the resource cost pools differ in size and when there are large differences in the proportional resource usage of each cost pool (Labro & Vanhoucke, 2008).
DBC cost information is attached to DBC categories by the use of a series of cost allocation procedures. The first is the allocation of the costs of support cost centers – e.g., personnel, communications, finance and security – to final cost centers, like clinical departments. This has mostly been done using rather simple direct costing allocation rules (Zuurbier, 2004; Zuurbier & Krabbe-Alkemade, 2007). The Dutch Health Authority determines the cost of 4,500 hospital services from the weighted average across 15 to 25 'frontrunner' hospitals. Total hospital services are assigned to 15 resource-use categories. Total DBC costs are finally determined by the number of services used, based on weighting statistics, of which time is a relatively important factor (Tan, Ineveld, Redekop & Van Roijen, 2011). The DBC costing procedure has distinctive differences from the costing procedure followed in most DRG systems. Generally, hospital costs are allocated to DRGs on the basis of length-of-stay as a proxy for costs (Quentin, Geissler, Scheller-Kreinsen & Busse, 2011). This may lead to artificial homogeneity of DRG groups: they may contain diagnoses of similar length-of-stays that in fact use hospital resources in different amounts.
In our study, we do not have the necessary informati-

on to accurately estimate "true" DBC costs. This makes it impossible to assess measurement and specification errors. We therefore focus on aggregation errors by looking at the cost homogeneity of the DBC system. We use four indicators that are common measures of the quality of case mix systems. The first indicator represents the model's *cost-effectiveness* (or efficiency) by looking at the added information value of additional cost categories to the model. We further use two indicators of the *within-group homogeneity* of case mix classes: the average percentage of outlier cases in the case mix groups, and the average coefficient of variation of all groups. The fourth indicator is the Reduction in Variance (RIV) and measures the *predictive validity* of the case mix system (Palmer & Reid, 2001; Reid, Palmer & Aisbett, 2000). From the cost accounting literature, we may infer that the use of more resource pools, activity pools and allocation drivers leads to a more fine-grained cost system with more homogeneous product categories that attains a higher predictive validity of health care costs. However, a case mix system consisting of a relatively high number of product categories also runs a higher risk of including product categories with only a limited number of cases. A low-volume category is expected to have low homogeneity due to statistical noise caused by sampling variation (Reid, Palmer & Aisbett, 2000).

For the case mix system as a whole, we therefore have two partially contradicting expectations. A more fine-grained case mix system like the DBC system is expected to have more cost homogeneous categories and a higher predictive validity than a more aggregated, coarse-grained case mix system, like the DRG system. However, the more fine-grained DBC system may also compound measurement, aggregation and specification errors. It's relatively large number of low-volume categories may also suffer from sampling variation noise. This may cause fine-grained case mix systems to be less homogeneous and consequently have a lower predictive validity than more aggregated case mix systems.

3 Research method

The dataset used contains all DBCs registered and invoiced by Dutch hospitals in 2007. In order to obtain cost data, we linked the median unit cost of the care activities to the care profile of a unique identification number. Thus, the cost of the care profile is the sum of the cost prices of the care activities, which belonged to the same hospital, hospital location and DBC with a unique identification number. This leads to a reconstruction of total hospital costs of each registered DBC, and thus to total hospital costs when adding up all invoiced DBCs. The salaries of the medical specialists are not included in the total cost figures.

We compare different case mix systems with varying levels of granularity by breaking down the DBC dataset in different ways. The finest granularity is reached when DBC-level information is used. Our dataset contains 44,128 different unique DBC codes. The lowest granularity level is reached when codes are grouped on the diagnosis level, which leads to 2,339 different diagnose categories in our sample data. The diagnosis grouping may be considered to have an equivalent granularity level to most other DRG systems. One should, however, bear in mind that the DBC diagnoses lack uniformity and are not as well structured as in DRG systems, because medical specialties use their own coding lists; diagnoses are based on the CvZ80 list (a classification of diseases). Two additional alternative granularity levels can be reached by combining diagnoses with treatment groups (14,991 unique codes) and by combining diagnoses with type of care groups (6,432 unique codes). Treatment groups include specialism specific types of outpatient, daycare and inpatient treatments. Type of care groups define whether the activity is a regular treatment, a follow-up treatment or a peer professional consultation. For each alternative breakdown of the data, we calculate the quality scores for the aggregation level and the corresponding codes. The quality scores focus on three dimensions: (1) cost effectiveness; (2) within-group homogeneity; and (3) predictive validity of the system (Palmer & Reid, 2001). We measure cost effectiveness by counting the number of unique codes responsible for 80% of the cases or hospital costs. Cost-effective case mix systems are supposed to contain groups defined in such a way that each group represents a significant portion of cases or total hospital costs. For example, if each group represented an equal amount of hospital costs, then the cost effectiveness measure would be 80%. That is, 80% of the costs are represented by 80% unique case mix codes. For differences in cost effectiveness, we contrast the DBC-level and diagnosis-level data (simulating DRG-type systems).

The within-group homogeneity is measured in two ways. The first measure is the average percentage of outlier cases in the case mix groups. To identify outliers, we use the inter quartile range method and take as upper trim point: 1st Quartile + 1.5 * (3rd Quartile -1st Quartile) (Reid, Palmer & Aisbett, 2000; Palmer & Reid, 2001). The second measure is the Coefficient of Variation, which is:

$$CV = \frac{sd\ cost}{average\ cost}$$

A high coefficient of variation indicates a low within-group homogeneity. The theoretical maximum value for the CV of a normal distribution is 1. Thus, for trimmed data, where the departure from normality should not be substantial, CVs larger than 1 indicate poor within-group homogeneity.

The predictive validity of a case mix system is the degree to which total variance can be explained by the variance of group means around the population mean (between-group variance). The ratio of the between-group variance to the total variance is the reduction in variance (RIV) due to the variance between groups, as opposed to variance within groups. Consequently, the RIV measures the reduction in cost variation by the classification system used (Bland, 2000; Benton et al., 1998). The definition of RIV is:

$$RIV = \frac{\sum_{j=1}^{k}\sum_{i=1}^{n}(c_{ij} - \mu)^2 - \sum_{j=1}^{k}\sum_{i=1}^{n}(c_{ij} - \mu_j)^2}{\sum_{j=1}^{k}\sum_{i=1}^{n}(c_{ij} - \mu)^2}$$

Where c_{ij} is the cost of case i in group (aggregation level) j, μ_j is the average cost of the cases within group j, μ is the average cost of all the cases, n is the number of cases and k is the number of groups.

4 DBC data

The dataset used for this study is derived from the national *DBC Information System* (DIS data). The DIS contains production data of all Dutch hospitals, including 87 general hospitals and 8 university hospitals. All DBCs completed and validated in 2007 are included in our analysis.

The DIS data system was created in 2005, and our dataset is the third edition. In 2007, many hospitals were still in a process of fine-tuning the data registration process. We therefore expected the 2007 dataset to contain some errors. Subsequently, we checked and cleaned the dataset before using it for analysis. The number of DBCs in 2007 before cleaning the data was 14,950,930. We excluded 'empty' DBCs from our sample, which are DBCs without matching care activities. The percentage of 'empty' DBCs is 12.05% of all sample DBCs, evenly distributed across hospitals, time and diagnoses. DBCs with more than 100 activities attached and DBCs with negative costs are also excluded. Furthermore, DBCs of two specialty hospitals are excluded from the database, since we expect this production to be significantly distinct compared to the care activities of general and university hospitals. After the data-cleaning procedure, the database contains 12,477,934 DBCs registered by 93 hospitals.

Table 1 provides descriptive statistics of DBC production numbers and hospital costs in university hospitals and general hospitals, classified into large, medium and small hospitals. Small hospitals are hospitals with a budget below € 60 million. Average hospitals have a budget between € 60 million and € 120 million, while large hospitals operate on budgets larger than € 120 million. The total annual budget of the sample hospitals constitutes 39% of the annual budget of all Dutch hospitals. This proportion varies from 21% for university hospitals to 49% for medium hospitals. The main reasons for the difference is that the sample cost data does not include the cost of salaried specialists, the cost of special treatments and university hospital costs for teaching, training and research. The hospitals completed and charged over 12.4 million DBCs, with an average of 134,171 DBCs per hospital. The hospitals completed 44,128 unique DBC codes (i.e. the DBC codes that were registered at least once) of 2,339 unique diagnoses. On average, the episode-oriented DBC system includes 18.7 times more different codes than if the system was based on diagnoses.

From the total number of unique DBC codes registered we infer that university hospitals, large and medium-sized hospitals provide comparable wide ranges of medical services. Only small hospitals offer a significantly lower number of unique DBC-codes. University hospitals differ from general hospitals in the avera-

Table 1 Descriptive statistics and health care production (DBCs) of all Dutch Hospitals in 2007

Characteristics	All	University	Large	Medium	Small
Number of hospitals	93	8	27	37	21
Total annual budget (in million €)	€ 11,882	3,199	4,730	2,999	953
Average annual budget (in million €)	€ 128	399	175	81	45
Total number of completed DBCs	12,477,934	1,208,258	5,442,941	4,410,435	1,416,300
Average number of completed DBC per hospital	134,171	151,032	201,590	119,201	67,443
SD of average number of completed DBCs	64,022	27,418	63,331	31,116	17,175
Total number of unique DBC codes registered	44,128	27,990	33,437	29,395	19,079
Total number of unique diagnoses registered	2,339	2,193	2,265	2,040	1,841
Total DBC costs (in million €)	€ 4,590	687	2,006	1,455	442
Total DBC costs as percentage of total annual budget	39%	21%	42%	49%	46%
Average DBC costs (Total costs / total number DBCs)	€ 368	569	369	330	312

ge costs per DBC. University hospital DBC cost is, on average, 70% higher than the average DBC cost of small general hospitals.

5 Results

Cost-effectiveness

Table 2 presents the number and percentage of unique DBCs and unique diagnoses covering 80% of total cases and 80% of total costs. This table demonstrates that only a small number of codes captures the majority of cases and costs: 4% of DBC-codes and 14% of diagnose-codes represent 80% of all cases, and 3% of DBC-codes and 14% of diagnose-codes represent 80% of total costs.

A total number of 1533 DBCs (3%) and 321 (14%) diagnose-codes explains 80% of total costs: 4.7 times more DBC-codes than diagnose-codes. Table 2 also shows that the complexity of the DBC-system can be reduced significantly when focusing on the 80%-cost category: the DBC-diagnosis ratio can be reduced from 18.9 to 4.7.

There is not much difference between the numbers of DBCs and diagnose-codes explaining 80% of the cases and of total costs between the different general hospital groups, but there are differences between general hospitals and university hospitals. University hospitals have significantly more DBCs contributing to the 80% cases and costs. This indicates that the cases and costs of university hospitals are distributed over more different DBC codes than those in general hospitals. DBC-coding seems to pick up this difference better than do diagnostic codes, which means that most of these differences are more related to treatment than to diagnosis.

Figure 1 shows the number of unique DBC codes and diagnoses in relation to the cumulative total production costs of all sample hospitals. The codes are sorted according to decreasing marginal cost coverage. The curves clearly show that a relatively small proportion of DBCs and diagnoses represents a large share of total costs: 10% of unique DBC codes cover approximately 92% of the total costs of DBCs, while 10% of the unique diagnoses cover 74% of total costs. DBC categories beyond the first 10%-group show low and rapidly declining marginal cost coverage, whereas the marginal cost explanation of diagnoses-groups beyond the first 10% is significantly higher and decline at a lower rate. The DBC-curve shows that 85% of all DBC codes explain only 4% of total costs, while 60% of diagnoses-based codes explain the same proportion of costs. The DBC system's complexity does not seem to be very cost-effective: a relatively high number of cases explain only a small proportion of costs. The granularity of the two remaining groups (diagnoses and treatments, and diagnoses and care type) lies between those of the DBC- and diagnosis-based systems, and so do the respective cost coverage functions in Figure 1.

Figure 1 **Cumulative costs as a function of the proportion of unique DBCs and unique diagnoses**

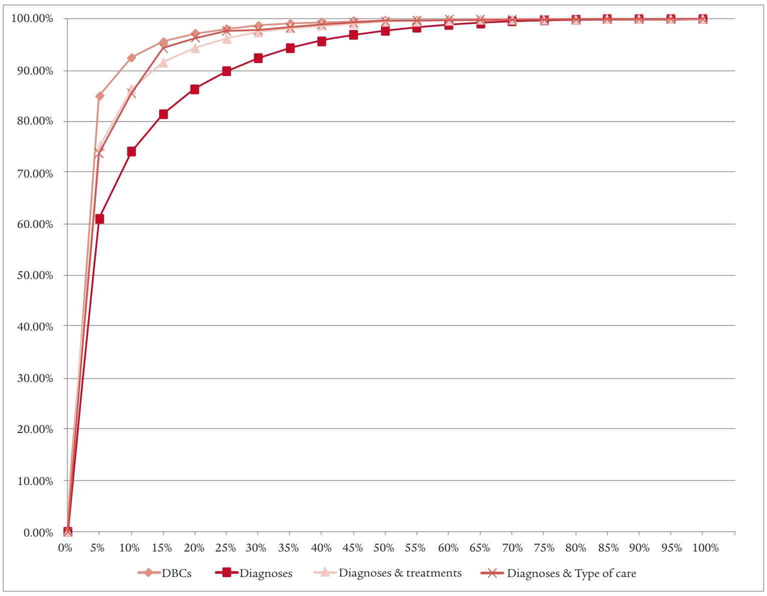

Table 2 **Number and percentage of unique DBCs and unique diagnoses generating 80% of total cases and 80% of total costs (for total sample and per hospital category)**

Type of hospital	Number (%) of DBCs				Number (%) of diagnoses			
	80% of all cases		80% of total costs		80% of all cases		80% of total costs	
	n	%	n	%	n	%	n	%
All	1730	4%	1533	3%	328	14%	321	14%
University	2792	10%	1886	7%	462	21%	410	19%
Large	1517	5%	1252	4%	304	13%	272	12%
Medium	1438	5%	1168	4%	288	14%	255	13%
Small	1292	7%	1019	5%	264	14%	242	13%

In small, medium and large hospitals, only 4 diagnoses appear to be responsible for 10% of the total costs of DBCs. These diagnoses are *chronic hemodialysis, supporting parturition including after care, basis care newborn babies* and *osteoarthritis (arthroplasty)*. In university hospitals, about 25 different diagnoses generate 10% of the total costs of DBCs.
Although we did not find many differences in number of DBC-codes and diagnoses explaining 80% of total costs between general hospital categories, this seems to be mainly an aggregation result. More differences exist between specializations than between hospital types, as Table 3 demonstrates. The variation in percentage of DBCs explaining 80% of total costs ranges from 3.2% for Orthopedics to 35% for Rehabilitation medicine.

Figure 2 shows the average number of DBC types per diagnosis for a specialty. The relationship between number of diagnoses and number of DBCs varies between specializations. This figure shows the diversity in the refinement of the DBC system, which is a result of the fact that the medical professions independently developed the DBC system. For example, the average number of DBC types per diagnosis is low for thoracic surgery, rehabilitation medicine and neurosurgery, but extremely high for urology. A reason for this difference is that urology used the "type of care" category to include a unique difference between health problems, e.g. stomachache, incontinence or infertility.

Table 4 presents the distribution of DBC costs, cases and number of unique DBCs over the three treatment settings: outpatient, daycare, and inpatient care. For 14 of the 21 specialties, it appears that most of the costs occur in the inpatient setting. Ophthalmology has the highest percentage of all specialties (33%) in daycare. Table 4 also shows that relatively more DBCs have been developed for outpatient settings than inpatient settings. It is quite remarkable that outpatient DBCs have driven the refinement of the DBC system, while most of the costs is in the inpatient setting.

Table 3 **Number of unique DBCs and diagnoses (% of total DBCs and diagnoses) generating 80% of total costs**

Specialty	Number of DBCs Generating 80% of costs	Total DBCs	Percentage	Number of diagnoses generating 80% of costs	Total diagnoses	Percentage
Allergology	67	323	20.7%	14	43	32.6%
Anesthesiology	56	537	10.4%	8	33	24.2%
Cardiology	45	456	9.9%	10	40	25.0%
Clinical genetics	28	64	43.8%	4	21	19.0%
Dermatology	83	652	12.7%	12	29	41.4%
Gastroenterology	244	5183	4.7%	25	74	33.8%
Geriatrics	26	565	4.6%	1	35	2.9%
Gynaecology	40	646	6.2%	9	48	18.8%
Internal medicine	246	3493	7.0%	46	283	16.3%
Neurology	99	2016	4.9%	23	110	20.9%
Neurosurgery	34	779	4.4%	73	113	64.6%
Opthalmology	99	1685	5.9%	23	73	31.5%
Orthopaedics	90	2829	3.2%	28	216	13.0%
Otolaryngology	70	1012	6.9%	12	58	20.7%
Paediatrics	186	2710	6.9%	47	291	16.2%
Plastic surgery	264	3823	6.9%	52	184	28.3%
Pneumonology	45	858	5.2%	9	50	18.0%
Psychiatry	29	277	10.5%	7	19	36.8%
Radiology	31	148	20.9%	15	60	25.0%
Radiotherapy	47	268	17.5%	5	14	35.7%
Rehabilitation medicine	70	200	35.0%	11	39	28.2%
Rheumatology	93	823	11.3%	16	81	19.8%
Surgery	171	3344	5.1%	54	184	29.3%
Thoraric surgery	39	573	6.8%	22	132	16.7%
Urology	674	10862	6.2%	16	74	21.6%

Figure 2 **Average number of DBC types per diagnosis (specialization level)**

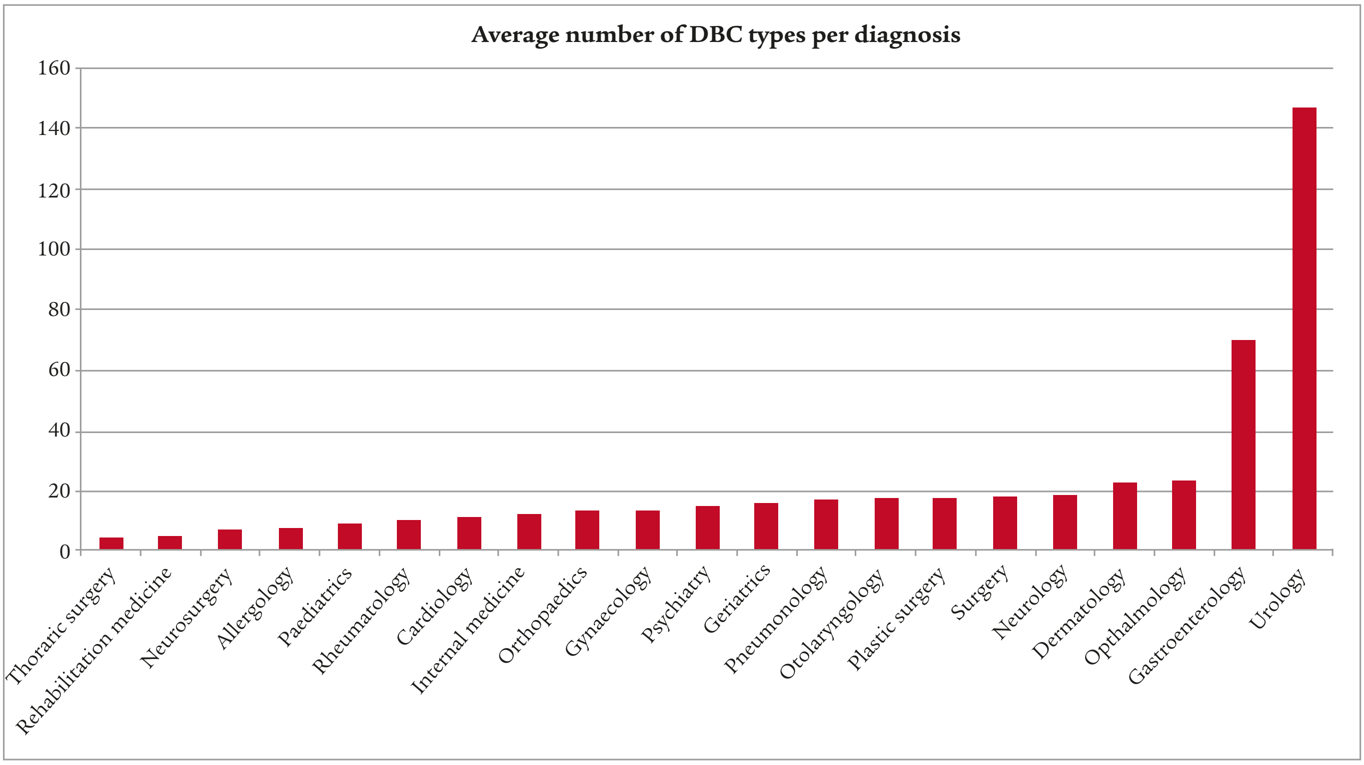

Table 4 **Percentage of total costs and number of unique DBCs according to treatment settings: outpatient care, daycare and inpatient care**

Specialty	Distribution of costs			Distribution of cases			Distribution of DBC codes		
	outpatient	daycare	inpatient	outpatient	daycare	inpatient	outpatient	daycare	inpatient
Allergology	0.84	0.16		0.94	0.06		0.74	0.24	0.02
Cardiology	0.14	0.05	0.81	0.74	0.06	0.20	0.47	0.23	0.30
Dermatology	0.78	0.11	0.11	0.98	0.02		0.76	0.15	0.09
Gastroenterology	0.10	0.31	0.59	0.59	0.27	0.14	0.52	0.28	0.20
Geriatrics	0.04	0.08	0.88	0.50	0.30	0.20	0.51	0.23	0.26
Gynaecology	0.19	0.08	0.73	0.71	0.09	0.20	0.31	0.25	0.44
Internal medicine	0.33	0.10	0.57	0.81	0.06	0.13	0.45	0.25	0.30
Neurology	0.19	0.05	0.76	0.83	0.05	0.12	0.57	0.20	0.23
Neurosurgery	0.05	0.03	0.92	0.58	0.09	0.33	0.59	0.07	0.34
Opthalmology	0.59	0.33	0.08	0.89	0.11	0.01	0.67	0.14	0.19
Orthopaedics	0.14	0.08	0.79	0.76	0.11	0.13	0.50	0.19	0.31
Otolaryngology	0.41	0.21	0.38	0.80	0.13	0.06	0.49	0.15	0.36
Paediatrics	0.13		0.87	0.72		0.28	0.50		0.50
Plastic surgery	0.18	0.26	0.56	0.65	0.23	0.12	0.41	0.25	0.34
Pneumonology	0.17	0.04	0.79	0.77	0.03	0.19	0.49	0.18	0.33
Psychiatry	1.00			1.00			1.00		
Rehabilitation medicine	0.79		0.21	0.86		0.14	0.68		0.32
Rheumatology	0.55	0.14	0.32	0.96	0.02	0.02	0.58	0.22	0.20
Surgery	0.17	0.06	0.77	0.80	0.07	0.14	0.47	0.18	0.35
Thoraric surgery			1.00	0.08		0.92	0.16		0.84
Urology	0.27	0.08	0.65	0.80	0.08	0.13	0.48	0.20	0.32

We may conclude that the cost-effectiveness of the DBC system is not very high: a large portion of the DBC codes (85%) explains only 4% of total costs. The marginal cost coverage of DBCs beyond the 10% group with highest cost representation is low and decreases rapidly. Because each of the medical specialties developed its part of the DBC system, most of the variation in cost-effectiveness is specialty-specific. The DBC system appears to be most fine-grained in the outpatient and daycare activities, which have the most low-cost patient settings. The high granularity of the system has not been applied in the inpatient setting, which is the setting in which a higher granularity in costs would have been more beneficial.
For subsequent analysis, we excluded DBCs with less than 30 registered cases. Low numbers of invoiced DBCs may lead to instability in costs attached across hospitals and over time because of sampling variation noise (Palmer & Reid, 2001). The DBC system appears to have many low-volume groups: 30,964 DBC codes of the total 44,128 were excluded, representing 70% of all unique DBC codes. The large number of unique DBC codes excluded represents only 186,936 cases, which is 1.5% of total cases (refer to Table 5). This is also an indication of the DBC system's low cost-effectiveness.
The remaining DBCs were further checked based on the existence of high-cost outliers. For this purpose, we applied the inter-quartile method commonly used in other studies (Kulinskaya, Kornbrot & Gao, 2005; DBC-Onderhoud, 2007a). This procedure is followed for each of the four alternative data aggregation levels separately. This led to a further exclusion of 11.4%, 11.7%, 15.6% and 16.5% of the cases, respectively.

It is generally believed that, when a case-mix system is derived from data of reasonable quality, no more than 5% of all cases will fall in the high outlier category (Palmer & Reid, 2001). The 2007 DBC system clearly does not reach this standard. The high exclusion percentages are not related to the cost system's granularity level. An alternative explanation may therefore be the quality of the 2007 dataset used. We do not expect the 2007 dataset to be of exceptionally poor quality, given the fact that 2007 was the third year of using the DBC system. However, the complexity of the system, combined with the administrative difficulties that were reported in the early years of the DBC system's existence, give some reason to expect that some measurement errors may have occurred.

Within-group homogeneity
The homogeneity *within* case mix groups (CV) is tested on four different classification systems, arranged in order of decreasing granularity: cost systems based on DBCs, on diagnoses and treatments, on diagnoses and type of care, and on diagnoses. A CV value higher than one is generally considered to indicate poor within-group homogeneity . The most fine-grained DBC classification turns out to be also the most homogeneous system. As the classifications become coarser, CV values rise to 0.70 for the diagnoses-based system. The average CV value of all classifications shows an acceptable within-group homogeneity in each classification, while only a small percentage of codes in groups 3 and 4 signal poor within-group homogeneity (13% and 19% of all codes, respectively). Although the DBC classification seems to outperform all other categorizations, every system reaches an acceptable level of within-group homogeneity.

The CV-value and average DBC cost in the best performing DBC classification system turns out to be posi-

Table 5 Number of unique groups and cases in different classification systems (after data cleaning)

Data cleaning procedures	Before cleaning		After cleaning		Reduction in %	
	Nr of groups	Nr of cases	Nr of groups	Nr of cases	Nr of groups	Nr of cases
Removal of codes with less than 30 cases						
Group1: based on DBCs	44,128	12,477,934	13,164	12,290,998	70.17%	1.50%
Group 2: based on Diagnoses and treatments	14,991	12,477,934	8.089	12,419,679	46.04%	0.47%
Group 3: based on Diagnoses and type of care	6,431	12,477,934	4,189	12,477,410	34.86%	0.15%
Group 4: based on Diagnoses	2,339	12,477,934	2,057	12,477,896	12.06%	0.02%
Removal of outliers using inter-quartile method						
Group1: based on DBCs	13,164	12,290,998	13,164	10,893,367	0.00%	11.37%
Group 2: based on Diagnoses and treatments	8,089	12,419,679	8,089	10,968,444	0.00%	11.68%
Group 3: based on Diagnoses and type of care	4,189	12,459,410	4,189	12,515,879	0.00%	15.60%
Group 4: based on Diagnoses	2,057	12,474,896	2,057	12,416,252	0.00%	16.50%

Table 6 Coefficient of variation for different classification groups

Classification group	CV<0.5	CV 0.5-1.0	CV>1.0	Average CV value
Group 1: based on DBCs	67.55%	31.24%	1.21%	0.36
Group 2: based on Diagnoses and treatments	59.41%	39.25%	1.34%	0.42
Group 3: based on Diagnoses and type of care	46.43%	40.92%	12.65%	0.55
Group 4: based on Diagnoses	31.36%	49.15%	19.49%	0.70

tively related. The lowest CV-values are found in the low-cost DBCs, which are mostly the outpatient and daycare treatments. The granularity in these groups is highest, whereas in the 10% most-expensive clinical treatments, the average CV value is 0.8. This is still acceptable, but the DBC system clearly does not focus on the code groups that could benefit most from a higher granularity level.
Figure 3 presents the relationship between CV value and the average costs for the DBC classification and the diagnosis classification for cases with average costs of between 10 and 800 euro.

The DBC classification shows the highest within-group homogeneity, whereas the diagnosis classification shows the highest variation in both low-cost and high-cost cases. Almost all DBCs fall under the 1 threshold, while a significant number of diagnosis-based codes appear to have a CV-value larger than 1.

Predictive validity
The system's predictive validity is measured by the reduction in variance factor. The different alternative classifications lead to significant differences in predictive validity scores (see Table 7).

The DBC system has the highest predictive validity and the diagnoses-based system has the lowest, with the other two systems showing RIV-values in between. The difference between the RIV scores of the first two classification groups is minimal, which means that a reduction of the 13,164 DBC codes to 8,089 diagnoses/treatment combinations (reducing the number of codes by 39%) does not reduce the system's predictive validity. The main reason for this result is the inclusion of the type of care and care demand categories in the DBC system. These categories do not lead to a proportionate reduction of the variation in costs. Also, the treatment category contributes most to the reduction of cost variance. This also becomes evident when the results of the alternative classifications in the DBC system are compared with the predictive validity scores of other existing DRG systems (see Table 8).

The Dutch DBC system has a relatively high RIV compared with other case mix systems in use elsewhere. The high score, however, is not only caused by the high

Figure 3 Scatter plot for relation between average costs and CV value for DBCs and diagnoses

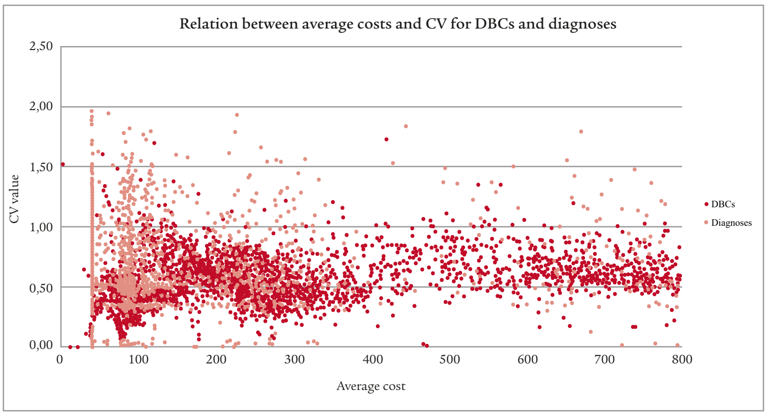

Table 7 Predictive validity of different classification systems

Classification group	RIV
Group 1: based on DBCs	0.664
Group 2: based on Diagnoses and treatments	0.662
Group 3: based on Diagnoses and type of care	0.520
Group 4: based on Diagnoses	0.483

Table 8 Examples of predictive performance of different international DRG systems and the Dutch DBC system

Paper	Classification system	RIV
Freeman 1991	DRG Barcelona hospitals	0.336
	Refined DRGs Barcelona hospitals	0.353
	MEDPAR sample 1996	0.214
	OHIO database 1986	0.307
Freeman 1995	DRG refinement model	0.381
Averill 1995	US HCFA DRGs	0.315
	AP DRGs	0.362
	R-DRGs	0.364
	APR-DRGs	0.410
Dutch DBC 2007 (this study)	DBCs (excluding low volume groups & excluding outliers)	0.664
	DBCs (including low volume groups & including outliers)	0.424

granularity of the system but also by the exclusion of 70% of the DBC-codes, representing 13% of the cases.

6 Conclusion and discussion

Cost accounting literature suggests that, in general, more fine-grained cost accounting systems with a larger number of cost pools and allocation rates produce more accurate cost information. However, under certain conditions, a more fine-grained cost accounting system may not lead to more accurate cost information. For instance when an erroneous classification system is used, when the complexity of the system leads to measurement errors, or when in a considerable number of categories only a few observations are registered. A more coarse-grained costing system may produce more reliable cost information, provided that measurement errors, aggregation errors and specification errors partially offset each other in the aggregation process. Our database does not provide the tools to assess the absolute accuracy of DBC cost information, because an error-free benchmark cost for each of the DBC case mix groups is not available. We therefore compared four case mix systems with different aggregation levels on four performance criteria commonly used in assessing case mix systems' information quality. The classifications used are 1) the fine-grained DBC system with 44,128 different codes, 2) the diagnoses and treatment classification with 14,991 codes, 3) the diagnoses and type of care classification with 6,431, and 4) the relatively coarse-grained classification based on diagnoses with 2,339 different case mix codes. A remarkable finding is that the different datacleaning operations resulted in the exclusion of many cases (13%, mostly due to a high number of outliers) and of many codes (70%, mostly because of very few observations). The large number of outlier cases and low-frequency codes may be caused by measurement errors or specification errors. A fine-grained DBC-system generally requires from registrants to make a selection between categories, displaying subtle differences. This may lead to measurement errors evidenced by a large number of outliers. A fine-grained DBC-system may also include treatment categories that are very specific and specialized. This may lead to specification errors because of a large variation in treatment procedures. This is especially the case in DBC's with low number of observations and large sampling variation noise. Fine-grained case mix systems give rise to measurement and specification errors, while coarse-grained systems may partly offset these errors when activity cost pools (or case mix codes) are more aggregated.

The DBC system does not appear to be very cost-effective. A relatively high number of cases represent only a small proportion of costs: 85% of all DBC codes explain only 4 % of total costs, whereas some 60% of diagnose-based codes explain the same proportion of costs. A total number of 1533 DBCs (3% of total DBC codes) and 321 diagnose groups (14% of total diagnose codes) explain 80% of total costs. The DBC marginal cost explanation beyond the first 10% is significantly lower and declines more rapidly than the marginal cost explanation of the diagnose-based case mix groups. The level of cost-effectiveness differs between medical specializations, which is to be expected given the fact that the medical professions had a great influence on the design of the DBC system. A large number of DBC codes have been developed for outpatient settings, but a relatively low number of DBCs were developed for inpatient treatments. This is surprising, since both total costs and cost variation is expected to be highest in inpatient settings. This would also call for a relatively higher percentage of DBC codes to be inpatient codes.

The different classification methods produce expected results on cost homogeneity and predictive validity. The most fine-grained DBC classification reaches the highest within-group homogeneity and predictive validity, whereas the most coarse-grained diagnoses-based classification produces the lowest scores. The two alternative classifications reach intermediate results on both dimensions. Although the fine-grained DBC system scores best on both dimensions, this does not mean that the diagnose-based system's performance is unacceptable. It only reaches unacceptable levels of homogeneity in 19% of the codes, but it performs reasonably well on the average CV-value of 0.70. It also scores reasonably well on predictive performance, especially in comparison with other existing DRG systems.

The analysis of the performance of alternative classifications in the DBC system shows that using more fine-grained classification systems lead to improvement of cost homogeneity and predictive validity. The DBC system, because of its specific design qualities, reaches these improvements at the cost of excluding many cases and case mix groups. Furthermore, our analyses suggest the existence of measurement and specification errors in the DBC system. Using a fine-grained case mix system under these conditions may increase total error in product costing information, because of the compounding effect of the measurement and specification errors on total product cost. Finally, we found that most of the case mix groups were developed for outpatient activities, whereas most costs – and the highest cost variations – are found in inpatient settings. Using more fine-grained cost systems for inpatient case episodes and less case mix groups for outpatient settings would perhaps have led to improved cost information and a more cost-effective case mix system.

This study looked at the effect of alternative classification systems on the quality of cost information. The strength of this study is that a very fine-grained reallife system was used to simulate different levels of granularity of cost data. We used the same database controls for differences in system design, and measurement methods that would otherwise influence the results if different DRG systems were used. The weakness is the absence of an objective and fully accurate cost number for each case mix group that could have been used as a benchmark for the accuracy of the case mix systems. Instead, we used different alternative case mix system characteristics in order to reconstruct the systems' performance. Although the DBC system outperformed other systems on cost homogeneity and predictive validity, it did so in an inefficient way. This paper shows that similar levels of predictive validity can be reached by using diagnosis as the basis of the classification system and adding treatment types to come to more cost homogeneous categories. The current DOT system therefore seems a sensible alternative for the DBC system. ■

Dr. Yvonne Krabbe-Alkemade is senior beleidsmedewerker van de Nederlandse Zorgautoriteit.
Prof. dr. Tom Groot is hoogleraar Management Accounting aan de Vrije Universiteit.

References

- Benton, P.L., Evans, H., Light, S.M., Mountney, L.M., Sanderson, H.F., & Anthony, P. (1998). The development of Healthcare Resource Groups--Version 3. *Journal of Public Health Medicine, 20*(3): 351-358.
- Bland, J. (2000). An introduction to medical statistics. Oxford University Press.
- Busse, R., Geissler, A., Quentin, W., & Wiley, M. (2011). *Diagnosis-Related Groups in Europe; moving towards transparency, efficiency and quality in hospitals*. Maidenhead UK: McGraw-Hill Open University Press. Retieved from
- http://www.euro.who.int/__data/assets/pdf_file/0004/162265/e96538.pdf.
- Christensen, J., & Demski, J.S. (1995). The classical foundations of'modern'costing. *Management Accounting Research, 6*(1): 13-32.
- Cooper, R., & Kaplan, R.S. (1988). Measure costs right: make the right decisions. *Harvard business review, 66*(5): 96-103.
- Cooper, R., & Kaplan, R.S. (1992). Activity-based systems: Measuring the costs of resource usage. *Accounting Horizons, 6*(3): 1-13.
- Datar, S., & Gupta, M. (1994). Aggregation, Specification and Measurement Errors in Product Costing. *Accounting Review, 69*(4): 567-591.
- DBC-Onderhoud. (2007a). *DBCs eenvoudig beter: de kunst van het weglaten. Gezamenlijk plan van aanpak bestuurlijk overleg*. Utrecht.
- DBC-Onderhoud. (2007b). *DBCs op weg naar transparantie, deel III*. Utrecht.
- Fetter, R.B., Brand, D.A., & Gamache, D. (1991). *DRG's: Their design and development*. Ann Arbor, MI: Health Administration Press.
- Gupta, M. (1993). Heterogeneity issues in aggregated costing systems. *Journal of Management Accounting Research, 5*: 180-212.
- Jackson, T. (2000). Cost estimates for hospital inpatient care in Australia: evaluation of alternative sources. *Australian and New Zealand Journal of Public Health, 24*(3): 234-241.
- Kobel, C., Thuilliez, J., Bellanger, M., & Pfeiffer, K.-P. (2011). DRG systems and similar patient classification systems in Europe. In R. Busse, A. Geissler, W. Quentin, & M. Wiley (Eds.), *Diagnosis-Related Groups in Europe; Moving towards transparency, efficiency and quality in hospitals* (pp. 37-58). New York: McGraw-Hill.
- Kulinskaya, E., Kornbrot, D., & Gao, H. (2005). Length of stay as a performance indicator: robust statistical methodology. *IMA Journal of Management Mathematics, 16*(4): 369-381.
- Labro, E., & Vanhoucke, M. (2007). A simulation analysis of interactions among errors in costing systems. *Accounting Review, 82*(4): 939-962. doi: 10.2308/accr.2007.82.4.939.
- Labro, E., & Vanhoucke, M. (2008). Diversity in Resource Consumption Patterns and Robustness of Costing Systems to Errors. *Management Science, 54*(10): 1715-1730. doi: 10.1287/mnsc.1080.0885
- Nederlandse Zorgautoriteit (NZa) (2011). *Besluit productstructuur DOT. Letter to the Minister of Health, Welfare and Sport*.
- Palmer, G., & Reid, B. (2001). Evaluation of the performance of diagnosis-related groups and similar casemix systems: methodological issues. *Health Services Management Research, 14*(2): 71-81.
- Quentin, W., Geissler, A., Scheller-Kreinsen, D., & Busse, R. (2011). Understanding DRGs and DRG-based hospital payment in Europe. In R. Busse, A. Geissler, W. Quentin & M. Wiley (Eds.), *Diagnosis-Related Groups in Europe; Moving towards transparency, efficiency and quality in hospitals* (pp. 23-35). Maidenhead, Berkshire, UK: Open University Press.
- Reid, B., Palmer, G., & Aisbett, C. (2000). The performance of Australian DRGs. *Austalian Health Review, 23*(2): 20-31.
- Schreyogg, J., Stargardt, T., Tiemann, O., & Busse, R. (2006). Methods to determine reimbursement rates for diagnosis related groups (DRG): a comparison of nine European countries. *Health Care Management Science, 9*(3): 215-223.
- Steinbusch, P.J., Oostenbrink, J.B., Zuurbier, J.J., & Schaepkens, F.J. (2007). The risk of upcoding in casemix systems: a comparative study. *Health Policy, 81*(2-3): 289-299.
- Tan, S.S., Ineveld, M. van, Redekop, K., & Rojien. L.H. van (2011). *The Netherlands: The diagnose behandeling combinaties*. Maidenhead, Berkshire: Open University Press.
- Westerdijk, M., Zuurbier, J., Ludwig, M., & Prins, S. (2012). Defining care products to finance health care in the Netherlands. *European Journal of Health Economics, 13*(2): 203-221. doi: 10.1007/s10198-011-0302-6
- Zuurbier, J.J. (2004, december 2004). *Model Kostprijzen, DBC 2003, versie 17*. DBC Onderhoud, Utrecht.
- Zuurbier, J., & Krabbe-Alkemade, Y. (Ed.). (2007). *Onderhandelen over DBC's* (2 ed.). Maarssen: Elsevier Gezondheidszorg.