Testing the generalizability of national reimbursement rates with respect to local setting: the costs of abdominal aortic aneurysm surgery in Denmark

Sören Løvstad Christensen, Mette Kjøelby, Lars Ehlers

Health Technology Assessment and Health Services Research, Centre for Public Health, Central Denmark Region, Denmark; School of Public Health, Aarhus University, Denmark; Health Economics and Management, Aalborg University, Denmark

Objective: The purpose of this study is to investigate if the Danish national diagnosis-related group (DRG) tariffs for surgery for abdominal aortic aneurysm (AAA) were good estimates of the actual costs in two local hospitals in the Central Region of Denmark.

Methods: We collected clinical data for 178 AAA patients operated at Skejby Hospital and Viborg Hospital in the period 2005–2006 from the Danish National Vascular Registry and economic data from the administrative systems in the hospitals. We used bootstrap methods to calculate 95% confidence intervals (CIs) for the mean costs of surgery for ruptured AAA, nonruptured AAA and AAA where the patient died within 30 days by applying a cost-trimming rule that the Danish National Board of Health uses in calculating national DRG tariffs.


Conclusion: The bootstrap method is useful for testing the generalizability of national DRG tariffs as estimates of local surgical costs.

Keywords: bootstrap method, costs, DRG, abdominal aortic aneurysm

Introduction

National diagnosis-related group (DRG) tariffs are the basis for reimbursement of inpatient hospital costs in many countries.1–3 In Denmark, the DRG tariffs are recalculated each year by the Danish National Board of Health by applying a full-cost accounting principle to the Danish DRG case-mix based on detailed reports of costs and activities from 31 hospitals responsible for approximately 60% of hospital discharges in Denmark. The DRG tariffs are calculated as the mean costs for all activities in the respective DRGs and used as a central tool for reimbursement and decision making in the Danish health care system. Because of the financial implications, the validity and generalizability of the national DRG tariffs are often questioned, and continuous analyses of the appropriateness of the tariffs are warranted.1–4

The purpose of this study is to use the bootstrap method to perform tests of the generalizability of national DRG tariffs with respect to local setting. We chose the case of surgery for abdominal aortic aneurysms (AAAs) as an empirical example because the DRG tariffs for AAA surgery have been fluctuating since the introduction of the Danish DRG system in 2002, and a number of health economic...
studies in this field have questioned the appropriateness of using national DRG rates as proxies for surgical costs.5–7 We, therefore, performed tests to investigate if the Danish DRG reimbursement rates for surgery for AAA were a good estimate for the actual costs of AAA surgery in the Central Region of Denmark in 2005–2006.

**Materials and methods**

**Patient data**

Our dataset consisted of 178 male patients having experienced ruptured or nonruptured AAA in the period 2005–2006. No data on surgical technique were available. In 2006, endovascular AAA surgery was approximately 10% of all AAA surgery in Denmark. Mortality rates in both hospitals were similar to national averages for rupture and elective surgery. The mean age of these patients was 71.37 years. The descriptive statistics of the patient group is presented in Table 1.

The patients were operated at one of the two hospitals in the Central Region of Denmark with a cardiovascular surgery unit, Viborg Hospital (86/178) and Skejby Hospital (92/178). Of the 178 patients, 31 (17%) had experienced a rupture, 117 (66%) underwent surgery for a nonruptured aneurysm, and 30 (17%) were registered as dead. Patients were registered as dead if they had died within 30 days after surgery for AAA.

We obtained data for theatre usage and total length of stay from the Danish Vascular Surgery Registry. Data for hours in intensive care unit (ICU) were obtained from the statistical departments in the local hospitals.

**Cost data**

A microcosting approach4,8 was used to calculate the total average cost for surgery for ruptured AAA, nonruptured AAA, and death in close collaboration with the economic departments in the hospitals. Special attention was given to the risk of double-counting and omitting cost items. We calculated the costs per patient as the sum of the costs of theatre usage, costs of stay at the ICU, and costs of stay in the general ward. This has been shown to be the main cost drivers of surgery for AAA.5,10 Information on unit cost was calculated from the management accounting systems at the hospitals. Costs were calculated including overhead in 2006 prices.

**Calculations of the total average costs per patient**

The total average cost per patient was calculated as the sum of the following three main cost drivers, which adds up to the costs included in the DRG tariffs:

1. **Cost of theatre usage:** The total cost per patient was calculated as the sum of the labor costs (the average number of hours in theatre times and the average wage rate per hour for each participant in the theatre) and calculated overhead costs. The hourly labor cost was estimated for surgeons, nurses, and others as actual wages including pension assuming approximately 1,700 effective working hours per year. Assumptions used to estimate theatre staff in procedures were obtained through literature and interviews. The overhead costs were calculated according to the type of patient, ie, ruptured AAA, nonruptured AAA, and death. We excluded certain types of overhead costs (the hotel costs) in this calculation of the cost of theatre usage in order to avoid double-counting.

2. **Cost of stay at the ICU:** The total costs per patient were calculated as the cost per day times the number of days in ICU. The costs per day in ICU were obtained from the management accounting systems in the hospitals.

3. **Length of stay in general ward:** The total costs per patient were calculated as the cost per day times the number of days in general wards. The costs per day in general ward unit were obtained from the hospital management accounting systems.

**The Danish DRG system**

A new Danish case-mix system including DRGs for inpatient services was implemented in Denmark in 2002.11,12 The system is widely used for reimbursement and as a tool for analyzing costs and activities in the Danish health care sector.
Like any other case-mix system, hospital services are grouped into resource homogenous groups, and tariffs are calculated for each group to represent the average total cost of the services within the particular group. The Danish case-mix system consists of 599 DRG tariffs and 93 Danish Ambulatory Group System (DAGS) tariffs (2006 version). A DRG/DAGS tariff is defined as an average per diem or case-mix group cost for an activity belonging to a resource homogenous group. It is the intention of the case-mix system that each tariff should reflect the average costs of treating a typical patient belonging to the particular group. (DAGS is used for ambulatory patients defined as patients treated in an ambulatory department, whereas DRG is used for inpatient treatment defined as patients treated in a bed department). In principle, a DRG/DAGS tariff includes all hospital costs needed to perform an activity from this group, ie, both variable costs, such as labor and materials, and fixed overhead costs. However, depreciation and financial interests on buildings, civil servant pensions, and some research expenses are excluded.

The tariffs are updated by the National Board of Health each year based on detailed reports of costs and activities from the participating hospitals. The report produced by each hospital includes a step-wise allocation of all hospital costs to final cost centers (whose output can be linked to patient contacts). The costs at the nonclinical overhead departments are allocated to the other overhead departments and after that to the final cost centers. This allocation is based on national guidelines and entails fixed or prioritized allocation bases for overhead.

The Danish DRG system divides AAA patients into three different DRG groups: ruptured, nonruptured, and dead. The DRG tariffs in Danish Krone (DKK) for these three groups were 108,554 DKK (ruptured), 88,016 DKK (nonruptured) and 48,588 DKK (dead) in 2006.7

Tests
We performed two statistical tests of the hypothesis that our calculated costs equal the DRG rates. Due to the skewed nature of the cost data, we decided to apply simple bootstrap methods13,14 to construct 95% confidence intervals (CIs). This procedure has the advantage that we do not have to make any distributional assumptions, and it has been recommended as the primary statistical test for making inferences about arithmetic means for small-sized samples of skewed cost data.15 The bootstrap method is based on repeated sampling from the observed data to calculate nonparametric CIs (we used 1,000 replicates). We also performed a Student’s t-test based on 95% confidence limits for comparison purposes.

Finally, we applied the trimming rule from the Danish DRG system on our data and performed the same tests again. Trimming the data means changing the value of outliers to a certain maximum. In the Danish DRG system, outliers are defined as observations outside the 95% quartile and these observations are given the value of the 95% quartile. The trimming point is 20 bed days for ruptured AAA, 33 days for nonruptured AAA, and 1 day for death. All tests were carried out using Stata 9.0 (StataCorp, College Station, TX).

Results
The costs of theatre usage in the two hospitals are presented in Table 2. It is seen that the two hospitals differ a great deal with respect to overhead costs. This reflects the fact that Skejby Hospital is equipped with more sophisticated technology, ie, more capital intensive. This is also indicated in the costs per stay in ICU, where the costs for Skejby Hospital and Viborg Hospital were 16,037 DKK and 13,892 DKK per day, respectively. The costs per day in the general ward were 4,345 DKK and 3,414 DKK for Skejby Hospital and Viborg Hospital, respectively.

The results of the calculation of total average costs per patient are shown in Table 3. It is seen that there is a large difference between the mean and the medians for all groups; this merely illustrates that the cost data are right-skewed. The skewed data, caused by some extreme values, give rise to high standard deviations.

The results of the statistical test of whether the Danish DRG tariffs were a good estimate of the actual costs of surgery for AAA in the Central Region of Denmark are shown in Table 4.

From the bootstrap intervals, it is observed that three DRG tariffs lie within the 95% CI (CI ruptured AAA, 98,178–195,327 [€13,196–€26,254]; CI nonruptured AAA, 79,039–98,178 [€10,624–€13,196]; CI dead, 42,023–111,685 [€5,648–€15,011]). This means that we cannot reject equality between our estimates and the DRG tariffs. The t-statistics for the dead, nonruptured, and ruptured patients were 1.50, 0.11, and 1.52, respectively. Hence, with a significance level

<table>
<thead>
<tr>
<th>Table 2 Cost per hour of theatre usage (DKK per hour)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overhead cost</strong></td>
</tr>
<tr>
<td><strong>Skejby hospital</strong></td>
</tr>
<tr>
<td>Rupture</td>
</tr>
<tr>
<td>Nonrupture</td>
</tr>
<tr>
<td>Death</td>
</tr>
</tbody>
</table>
Table 3 Costs of surgery for abdominal aortic aneurysm (costs in DKK)

<table>
<thead>
<tr>
<th></th>
<th>Mean Nontrimmed</th>
<th>Mean Trimmed</th>
<th>Standard deviation Nontrimmed</th>
<th>Standard deviation Trimmed</th>
<th>Median Nontrimmed</th>
<th>Median Trimmed</th>
<th>Range Nontrimmed</th>
<th>Range Trimmed</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ruptured</td>
<td>146,717</td>
<td>138,557</td>
<td>138,940</td>
<td>105,116</td>
<td>116,008</td>
<td>116,008</td>
<td>753,025</td>
<td>500,060</td>
<td>31</td>
</tr>
<tr>
<td>Nonruptured</td>
<td>88,609</td>
<td>82,298</td>
<td>55,636</td>
<td>32,979</td>
<td>72,642</td>
<td>72,642</td>
<td>389,764</td>
<td>132,828</td>
<td>117</td>
</tr>
<tr>
<td>Dead</td>
<td>76,854</td>
<td>69,306</td>
<td>102,744</td>
<td>81,991</td>
<td>25,613</td>
<td>25,613</td>
<td>388,338</td>
<td>255,555</td>
<td>30</td>
</tr>
</tbody>
</table>

Table 4 Confidence intervals with significance level of 5% (costs in DKK)

<table>
<thead>
<tr>
<th></th>
<th>Lower bound Parametric</th>
<th>Lower bound Nonparametric</th>
<th>Upper bound Parametric</th>
<th>Upper bound Nonparametric</th>
<th>DRG</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ruptured</td>
<td>95,754</td>
<td>98,178</td>
<td>197,681</td>
<td>195,327</td>
<td>108,544</td>
</tr>
<tr>
<td>Nonruptured</td>
<td>78,421</td>
<td>79,039</td>
<td>98,796</td>
<td>98,178</td>
<td>88,016</td>
</tr>
<tr>
<td>Dead</td>
<td>38,488</td>
<td>42,023</td>
<td>115,219</td>
<td>111,685</td>
<td>48,588</td>
</tr>
<tr>
<td>Truncated</td>
<td>100,000</td>
<td>102,355</td>
<td>177,114</td>
<td>174,759</td>
<td>108,544</td>
</tr>
<tr>
<td></td>
<td>76,259</td>
<td>76,129</td>
<td>88,336</td>
<td>88,466</td>
<td>88,016</td>
</tr>
<tr>
<td></td>
<td>38,690</td>
<td>41,345</td>
<td>99,922</td>
<td>97,268</td>
<td>48,588</td>
</tr>
</tbody>
</table>

Abbreviation: DRG, diagnostic-related group.

of 5%, we could not reject equality between our estimates and the DRG rates.

Even though our aim was not to replicate the DRG rates, we must be aware of the fact that the Danish DRG system labels observations outside the 95% quartile as outliers and gives these observations the value of the 95% quartile. It is seen from Table 4 that by trimming our data in the same fashion as the Danish DRG system, we get slightly different CIs; however, our conclusions do not change because the DRG rates still fall within the CIs. By using Student’s t-test, we got t-statistics of 1.38 (dead), 1.58 (ruptured), and 1.87 (nonruptured), and hence this could not reject our hypothesis of equality. Since our conclusions do not change, our tests are robust to the effect that large values might have.

Discussion
Our objective was to use the bootstrap method to test whether the national DRG tariffs were good estimates of the real or observed average costs associated with treatment of AAA in two local Danish hospitals. We carried out simple tests based on the bootstrap method to investigate whether our estimates could be equal to the DRG tariffs, and our results showed that the estimates we get from our cost data could not be said to be different from the national DRG tariffs. Although the DRG tariffs have been fluctuating, the variation in tariffs in the period 2005–2007 has been within the bootstrap intervals. Hence, the DRG tariffs at first seem to be an appropriate measure of the cost associated with AAA.

However, before concluding on these results, three points must be taken into consideration. First, it is seen that the variation in the costs per operation is quite large. This is primarily due to patient-specific circumstances. The broad CIs should not be a result of relatively few observations in the sample if the empirical distribution of the sample data is an adequate representation of the true distribution of the costs of AAA surgery. This we cannot be sure of. Second, the national DRG tariffs for rupture and death do in several instances lie at the border of our CIs; however, changing the confidence level to 90% does not lead to rejection of equality between our estimates and the DRG estimates. Third, the costs differ at the two hospitals, and the conclusion from the tests might not apply to the hospitals alone.

We used bootstrap methods to make inferences about the mean of the skewed cost data as recommended by Desgagne et al.; and we applied to our sample the same trimming rules that have been used by the Danish National Board of Health in calculating the national DRG tariffs. Compared with other nonparametric tests of median costs, such as Wilcoxon style rank tests, the bootstrap method preserves the economically important characteristics of the data. We also applied normal Student’s t-test for comparison purposes; however, due to the skewness of our cost data and the relatively small amount of observations, the results derived from this procedure could be inferior to the bootstrap results. The Student’s t-test has advantages; however, its limitations are known to most readers.
As more countries adopt or adapt DRG-type case classification systems, decision makers and researchers are becoming increasingly reliant on national DRG tariffs. Yet, such tariffs do not necessarily reflect costs in different local settings, even when the clinical condition or procedure category appears similar. Furthermore, national DRG-type systems are not identical and are not utilized for the same purpose. Therefore, statistical tests of the generalizability of national reimbursement rates may be relevant to perform in many instances. We believe our study could serve as a relevant input or inspiration for decision makers and economic researchers in other settings who might want to investigate whether national reimbursement rates are good estimates of local costs.

Conclusion
The bootstrap method was applied to test the generalizability of national reimbursement rates with respect to local setting. Danish national DRG rates were found to be a good estimate of the costs of surgery for AAA in the Central Region of Denmark during the years 2005–2006.

Disclosure
The authors report no conflicts of interest in this work.

References