Journal of Medical Systems

Operating room efficiency before and after entrance in a benchmarking program for surgical process data

--Manuscript Draft--

Manuscript Number:	JOMS-D-17-00206R1
Full Title:	Operating room efficiency before and after entrance in a benchmarking program for surgical process data
Article Type:	Original Research Article
Section/Category:	Education & Training
Keywords:	Operating room; Utilization; Performance indicators; Benchmarking; Efficiency
Corresponding Author:	Vera Winter, MA, PhD Syddansk Universitet Odense, DENMARK
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Syddansk Universitet
Corresponding Author's Secondary Institution:	
First Author:	Sara Pedron, MSc
First Author Secondary Information:	
Order of Authors:	Sara Pedron, MSc
	Vera Winter, MA, PhD
	Eva-Maria Oppel, MSc, PhD
	Enno Bialas, MD
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	Background: Operating room (OR) efficiency continues to be a high priority for hospitals. In this context the concept of benchmarking has gained increasing importance as a means to improve OR performance. The aim of this study was to investigate whether and how participation in a benchmarking and reporting program for surgical process data was associated with a change in OR efficiency, measured through raw utilization, turnover times, and first-case tardiness. Materials and Methods: The main analysis is based on panel data from 202 surgical departments in German hospitals, which were derived from the largest database for surgical process data in Germany. Panel regression modelling was applied. Results: Results revealed no clear and univocal trend of participation in a benchmarking and reporting program for surgical process data. The largest trend was observed for first-case tardiness. In contrast to expectations, turnover times showed a generally increasing trend during participation. For raw utilization no clear and statistically significant trend could be evidenced. Subgroup analyses revealed differences in effects across different hospital types and department specialties. Conclusions: Participation in a benchmarking and reporting program and thus the availability of reliable, timely and detailed analysis tools to support the OR management seemed to be correlated especially with an increase in the timeliness of staff members regarding first-case starts. The increasing trend in turnover time revealed the absence of effective strategies to improve this aspect of OR efficiency in German hospitals and could have meaningful consequences for the medium- and long-run capacity planning in the OR.

Operating room efficiency before and after entrance in a benchmarking program for surgical process data

Sara Pedron, MSc

Institute for Health Economics and Health Care Management, Helmholtz Zentrum München, Germany. Ingolstädter Landstraße 1, 85764 Neuherberg (DE)

Vera Winter*, MA, PhD

Department of Political Science and Public Management, University of Southern Denmark, Denmark. Campusvej 55, 5230 Odense (DK)

Eva-Maria Oppel , MSc, PhD
Hamburg Center for Health Economics, Universität Hamburg, Germany. Esplanade 36, 20354 Hamburg (DE)

Enno Bialas, MD in Anesthesiology

digmed Datenmanagement im Gesundheitswesen GmbH. Flachsland 23, 22083 Hamburg (DE)

*corresponding author:

Vera Winter, MA, PhD

+45 6550 1884

winter@sam.sdu.dk

ABSTRACT

Background: Operating room (OR) efficiency continues to be a high priority for hospitals. In this context the concept of benchmarking has gained increasing importance as a means to improve OR performance. The aim of this study was to investigate whether and how participation in a benchmarking and reporting program for surgical process data was associated with a change in OR efficiency, measured through raw utilization, turnover times, and first-case tardiness.

Materials and Methods: The main analysis is based on panel data from 202 surgical departments in German hospitals, which were derived from the largest database for surgical process data in Germany. Panel regression modelling was applied.

Results: Results revealed no clear and univocal trend of participation in a benchmarking and reporting program for surgical process data. The largest trend was observed for first-case tardiness. In contrast to expectations, turnover times showed a generally increasing trend during participation. For raw utilization no clear and statistically significant trend could be evidenced. Subgroup analyses revealed differences in effects across different hospital types and department specialties.

Conclusions: Participation in a benchmarking and reporting program and thus the availability of reliable, timely and detailed analysis tools to support the OR management seemed to be correlated especially with an increase in the timeliness of staff members regarding first-case starts. The increasing trend in turnover time revealed the absence of effective strategies to improve this aspect of OR efficiency in German hospitals and could have meaningful consequences for the medium- and long-run capacity planning in the OR.

Keywords: Operating room, Utilization, Performance indicators, Benchmarking, Efficiency

1. Introduction

Health care providers are continuously striving to achieve higher levels of efficiency in response to health care challenges. Considerable attention within hospitals is dedicated to the operating room (OR) area, as one of the most important sources of hospital revenues and costs. Furthermore, OR process quality and safety is pivotal for the successful provision of hospital care [1,2]. As a result, increased attention in research and practice has been given to explore means to improve overall OR efficiency. In this context, the concept of benchmarking on standardized key performance indicators (KPIs) is increasingly applied to the health care industry in order to monitor OR efficiency and to improve performance [3-6]. Several benchmarking and reporting programs were developed in the last years in many European countries. The largest one in Germany, namely the *VOPM, BDA/DGAI, BDC benchmarking and reporting program*[1], was developed in 2010 and promoted by the main German professional associations of surgeons, anesthesiologists and OR managers. By entering the program, participants commit themselves to the monthly delivery of OR key process times and to a regular communication and update of structural hospital characteristics of surgical units (i.e. block time, allocated capacities, first case start target times), delivering also data from the year(s) previous to participation. The program is administered by an external IT provider, which ensures the compliance with current safety and privacy regulations and the plausibility of data. Participating hospitals have to pay an administrative participation fee per case to the IT provider.. The pooled data are then analyzed through several established and standardized key performance indicators (KPIs), which are made available on a web-based dashboard [3]. The main objective of this benchmarking and reporting program is to allow a reliable, objective and regular performance evaluation of the own perioperative processes and the identification of differences in OR efficiency among different facilities. As such participation itself has no direct effect on OR performance. However, external, reliable and detailed analysis tools could prove useful in targeting both critical aspects of low acceptance and poor use of analysis and controlling tools, in detecting

[1] Namely: VOPM (*Verband für OP-Management*, Association for OR management), BDA/DGAI *(Berufsverband Deutscher Anästhesisten*, Professional Association of German Anaesthesiologists & *Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin*, German Association of Anesthesiology and Intensive Care), BDC (*Berufsverbad der Deutschen Chirurgen*, Professional Association of German Surgeons).

inefficiencies, in increasing the accountability and involvement of staff members, all critical elements which are still a problem in many hospitals [1,6,7].

Prior research investigating the introduction and impact of KPIs as controlling and benchmarking tools focuses preponderantly on case studies at single institutions [1]. Thus, research exploring the impact of KPI benchmarking tools in a multiple institution setting and trends in OR efficiency in Germany remains scarce [8-10].

Addressing this research gap, the aim of this study is to investigate trends in OR efficiency of German hospitals participating in a benchmarking and reporting program. Specifically, we test whether and how hospitals differ in their OR efficiency before and after they enter a benchmarking program. In line with prior research and with the standard procedures issued by relevant stakeholders in the OR area, we define OR efficiency in terms of the following KPIs: raw utilization, turnover times, and first-case tardiness [3, 6, 10-16]. If OR efficiency improved after entering the program, this could indicate that participating in a benchmarking and reporting program helps to raise awareness and develop effective strategies to realize efficiency gains. Further, the results of our study enable detecting which areas might need further attention and development of the analysis tools. Finally, fruitful avenues for further research can be depicted.

2. Data and Methods

The study is a retrospective longitudinal evaluation of routinely documented data from the nursing/anesthesia database of each hospital, derived from the *VOPM, BDA/DGAI, BDC benchmarking and reporting program* for the period January 2009-July 2015. Previous studies have relied on this data already, corroborating its validity [17,18].

Data from the five most common surgical departments, namely general surgery (General S), trauma surgery (Trauma S), urology (Uro S), orthopedic surgery (Ortho S) and ear-nose-throat surgery (Ear-Nose S), were selected for the study as monthly average values for each department. Hospitals were then grouped into four hospital categories of similar size and complexity: general hospitals (General H, <5,000 cases/year), main hospitals (Main H, 5,000 -10,000 cases/year), maximum care hospitals (Max H, >10,000 cases/year), and university medical centers (Uni H, >10,000 cases/year and teaching/research activities). Furthermore, only hospitals with more than 50 beds and departments which operated on average at least four times a month were included.

For the present study, observations from one year before entrance to two years after entrance in the program were selected, leading to a total of 36 monthly observations for each department. For periods which dated back more than one year prior to entrance, operational data were scarcely available and the communication of structural characteristics (capacity allocations and planning agreements) was not accurate. By the time the research was started, most hospitals took part in the program for no longer than two years. Therefore, we limited the participation period considered in this study to two years. The database used for the main analysis is thus a semi-balanced panel data, which included hospitals which participated at least two years in the program by July 2015 and whose process data for one year before entrance were fully available. The semi-balanced structure results due to some missing monthly observations and from outlier correction[2].

To measure OR efficiency we focus on three KPIs, which are widely used in theory and practice as important metrics of OR efficiency [3, 6, 10-16]. First-case tardiness is defined as the difference between the unit specific start target time and the incision time of the first case of the day. Turnover time is defined as the time between suture of one case and incision of the next patient, i.e. as suture-incision-time. Raw utilization is defined as total incision-suture time as a percentage of total allocated capacity. It should be stressed that there exist several definitions of these KPIs. For example, turnover times are sometimes defined as the time where the OR remains empty between two patients [19,20]. In this paper we use the definitions currently in use in Germany, which focuses on the time between suture of one case and incision of the next patient, i.e., suture-incision-time as turnover time [3]. For a more detailed definition please see *Appendix A*."

To estimate our panel data models, we performed weighted fixed effect regression with clustering at the department level and heteroscedasticity robust standard errors. Based on the results of a robust Hausman test, we decided to perform weighted fixed effect regression modelling rather than random effects modelling [21]. In order to account for differences across hospital type and specialty, separate regressions for each group were carried out.

The main analysis was carried out estimating fixed effects at department level, considering each department as an independent unit. To ensure the integrity and robustness of our results, two sensitivity analyses were performed: First, to test whether results might be altered by unobserved variance at the hospital level, we re-ran our analyses using hospital fixed effects,

[2] Original observations at daily level which exceeded the lower and upper five percentiles were left out of monthly average calculations.

with clustering at the hospital level (*Online Resource 2: Appendix B*). Second, we re-ran our models using the full unbalanced dataset, including all hospitals participating in the program without the participation constraint of two years (unbalanced panel of 550 departments from 162 hospitals) (*Online Resource 2: Appendix C*).

The basic specification represented a static model with department specific fixed effects:

$$KPI_{i,t} = \beta_0 + \delta_0\, Year(1)_t + \delta_1\, Year(2)_t + \beta_1 ACD_{i,t} + \sum_{m=2}^{12} \gamma_m (Month_{i,t})_m + \alpha_i + u_{i,t}$$

Where:

i=1, 2,…202

t=-11, -10…+24

m=1,2, …12

The coefficients on the year dummy variables, *Year(1)* and *Year(2)*, indicate the difference in the level of KPI with respect to the year before entrance [*Year(0)*] and hence provide insights into the development of efficiency in the years after entering the program.

Average case duration (*ACD*) and month dummies (*Month*) were included as control variables. Prior research showed that *ACD* is positively correlated with case complexity [12,22], making the former a good proxy variable for otherwise unavailable information for risk adjustment. Furthermore, in German hospitals, efficiency is lower during summer or during months with several public or school holidays due to lower staff availability and flexibility. All analyses were performed using the Stata 14 statistical software.

3. Results

Average first-case tardiness before entrance was 11.31 (±8.64) minutes per case, significantly decreasing in the first two years of participation, according to the paired t-test (Table 1). The decrease can be witnessed for almost all hospital types and specialties for both years, yet with a quite large variation in magnitude. Average turnover time before entrance was 56.64 (±14.01) minutes per change, showing a statistically significant increasing trend after entrance in the program. Again, there is little variation in the direction of this trend across different hospital types and specialties, but some variation in magnitude. Average raw utilization before entrance was 47.03 (±12.91) percent. Overall, no significant differences between mean values for the consecutive years could be found. For most sub-analyses on hospital type and on specialty level, changes are insignificant as well.

According to the regression analysis, the first two years of participation in the benchmarking program were overall associated with a progressive and statistically significant decrease in average first-case tardiness (Tables 2-3). This trend led on average to a 1.3 minutes reduction of delays per case during the second year of participation compared with the year before participation.

Table 1: descriptive statistics and paired t-test results for the different hospital types and specialties considered.

	Late starts (min)					Turnover time (min)					Raw utilization (%)				
	Y(0)	**Δ Y(1) - Y(0)**		**Δ Y(2) - Y(1)**		**Y(0)**	**Δ Y(1) - Y(0)**		**Δ Y(2) - Y(1)**		**Y(0)**	**Δ Y(1) - Y(0)**		**Δ Y(2) - Y(1)**	
	Mean (SD)	**Diff (SE)**		**Diff (SE)**		**Mean (SD)**	**Diff (SE)**		**Diff (SE)**		**Mean (SD)**	**Diff (SE)**		**Diff (SE)**	
Overall	11.31 (±8.64)	-0.83 (0.25)	**	-1.31 (0.25)	**	56.54 (±14.01)	0.95 (0.41)	*	2.14 (0.41)	**	47.03 (±12.91)	-0.13 (0.36)		0.37 (0.36)	
Hospital type															
General H	10.69 (±8.5)	-0.5 (0.42)		0 (0.45)		49.13 (±9.1)	0.33 (0.47)		1.37 (0.48)	**	43.53 (±12.75)	0.2 (0.64)		0.69 (0.65)	
Main H	11.81 (±8.38)	-1.75 (0.46)	**	-2.29 (0.46)	**	53.23 (±10.69)	0.99 (0.58)		2.18 (0.6)	**	48.83 (±11.7)	-0.67 (0.62)		0.15 (0.62)	
Max H	9.12 (±7.28)	0.87 (0.51)		-0.68 (0.52)		57.52 (±9.34)	1.49 (0.64)	*	3.53 (0.68)	**	48.41 (±15.52)	-2.39 (0.91)	**	-1.37 (1.05)	
Uni H	12.81 (±9.47)	-1.39 (0.57)	*	-1.96 (0.56)	**	71.19 (±16.12)	0.03 (0.97)		0.12 (0.99)		48.58 (±11.43)	1.67 (0.69)	*	1.48 (0.65)	*
Specialty															
General S	11.26 (±7.79)	-0.93 (0.38)	*	-1.18 (0.39)	**	55.91 (±13.95)	1 (0.71)		2.36 (0.72)	**	50.94 (±12.42)	-0.42 (0.59)		0.61 (0.6)	
Trauma S	13.28 (±9.21)	-0.59 (0.56)		0.26 (0.56)		59.2 (±14.28)	1.31 (0.88)		3.33 (0.9)	**	47.02 (±13.12)	-0.51 (0.79)		-0.41 (0.81)	
Uro S	13.02 (±8.1)	-1.52 (0.55)	**	-2.3 (0.57)	**	59.92 (±12.94)	0.53 (0.87)		1.35 (0.88)		45.1 (±14.54)	-0.54 (0.9)		0.25 (0.97)	
Ortho S	9.91 (±8.5)	-0.27 (0.65)		-2.09 (0.62)	**	58.41 (±11.67)	1.65 (0.86)		2.45 (0.85)	**	45.45 (±11.1)	0.86 (0.83)		0.48 (0.8)	
Ear-Nose S	7.39 (±9.28)	-0.83 (0.75)		-2.29 (0.75)	**	48.17 (±13.79)	-0.05 (1.14)		0.36 (1.12)		42.33 (±10.76)	0.45 (0.86)		1.05 (0.85)	

Note: ** p<.01; * p<.05.

Hospital types: General H: general hospitals, Main H: main hospitals, Max H: maximum care hospitals, Uni H: university medical center.

Specialties: General S: general surgery, Trauma S: trauma surgery, Uro S: urology, Ortho S: orthopedic surgery, Ear-Nose S: ear-nose-throat surgery.

Similarly, a statistically significant and progressively decreasing trend was identified in both university medical centers (Uni H) and main hospitals (Main H). In these groups, benchmarking participation was associated with an average decrease of delays of more than 2 minutes per case during the second year. The trends in maximum care hospitals (Max H) and general hospitals (General H) were not statistically significant.

Table 2: fixed effects regression analysis results, overall and for the different hospital types.

Outcome		OVERALL		General H		Main H		Max H		Uni H	
Late starts (min)	**Year (1)**	-0.837	**	-0.395		-1.714	*	0.989		-1.587	**
		(0.32)		(0.377)		(0.701)		(0.837)		(0.552)	
	Year (2)	-1.276	**	0.046		-2.286	**	-0.143		-2.221	**
		(0.389)		(0.443)		(0.838)		(1.119)		(0.652)	
	Intercept	8.209	**	5.970	**	7.858	**	6.381	**	10.652	**
		(0.923)		(1.204)		(2.183)		(2.172)		(1.484)	
	adj. R^2	0.7532		0.8046		0.7486		0.5363		0.8143	
	N	7,252		2,268		2,052		1,240		1,692	
	N	202		63		57		35		47	
Turnover time (min)	**Year (1)**	0.774	**	0.471		0.844		1.841	**	0.158	
		(0.262)		(0.351)		(0.544)		(0.454)		(0.68)	
	Year (2)	1.925	**	1.528	**	1.902	*	3.855	**	0.864	
		(0.398)		(0.468)		(0.76)		(0.854)		(1.088)	
	Intercept	50.025	**	42.732	**	49.116	**	49.593	**	62.229	**
		(1.109)		(1.136)		(1.621)		(1.34)		(2.239)	
	adj. R^2	0.9013		0.8483		0.8455		0.8256		0.8775	
	N	7,272		2,268		2,052		1,260		1,692	
	N	202		63		57		35		47	
Raw utilization (%)	**Year (1)**	-0.040		0.175		-0.487		-2.612		1.964	*
		(0.472)		(0.579)		(0.787)		(1.748)		(0.856)	
	Year (2)	0.474		0.478		0.232		-1.562		1.957	
		(0.65)		(0.964)		(0.967)		(2.443)		(1.193)	
	Intercept	36.160	**	28.403	**	33.410	**	38.263	**	40.369	**
		(1.542)		(3.091)		(3.399)		(4.885)		(2.201)	
	adj. R^2	0.6688		0.7275		0.708		0.5293		0.647	
	n	7,186		2,250		2,024		1,238		1,674	
	N	202		63		57		35		47	

Note: ** p<.01; * p<.05. *Additional confounders:* average case duration (ACD), Months.

Hospital types: General H: general hospitals, Main H: main hospitals, Max H: maximum care hospitals, Uni H: university medical center.

Four of the five specialties included showed a generally decreasing trend in first-case tardiness in the first two years of participation. However, this decreasing trend was statistically significant only for three of the specialties (Ortho S, Uro S and Ear-Nose S) starting from the second year.

Results of the first sensitivity analysis using hospital specific fixed effects further supported the robustness of our findings. As shown in Appendix B, results remain stable in terms of direction, significance, and effect size (*Online Resource 2: Appendix B*).

Table 3: fixed effects regression analysis results, overall and for the different specialties included.

Outcome		OVERALL		General S		Trauma S		Ortho S		Uro S		Ear-Nose S	
Late starts (min)	Year (1)	-0.837	**	-0.942		-0.518		-0.199		-1.657		-0.907	
		(0.32)		(0.513)		(0.641)		(0.847)		(1.013)		(0.811)	
	Year (2)	-1.276	**	-1.271		0.665		-2.104	*	-2.625	*	-2.076	*
		(0.389)		(0.663)		(0.716)		(0.997)		(1.133)		(0.839)	
	Intercept	8.209	**	6.587	**	8.407	**	4.673	**	13.149	**	4.733	**
		(0.923)		(1.602)		(1.52)		(1.582)		(1.699)		(1.497)	
	adj. R^2	0.753		0.691		0.838		0.760		0.654		0.754	
	n	7,252		2,268		1,548		1,188		1,312		936	
	N	202		63		43		33		37		26	
Turnover time (min)	Y(1)	0.774	**	0.809		1.027		1.498	**	0.517		-0.359	
		(0.262)		(0.436)		(0.588)		(0.436)		(0.792)		(0.707)	
	Y(2)	1.925	**	2.123	**	2.956	**	2.072	**	1.362		0.214	
		(0.398)		(0.686)		(0.798)		(0.732)		(1.457)		(0.851)	
	Intercept	50.025	**	49.581	**	48.470	**	49.330	**	56.827	**	42.931	**
		(1.109)		(1.313)		(2.756)		(2.349)		(2.387)		(1.304)	
	adj. R^2	0.901		0.901		0.912		0.899		0.788		0.926	
	n	7,272		2,268		1,548		1,188		1,332		936	
	N	202		63		43		33		37		26	
Raw utilization (%)	Y(1)	-0.040		-0.159		-0.308		0.759		-0.490		0.253	
		(0.472)		(0.708)		(0.765)		(1.305)		(1.678)		(0.954)	
	Y(2)	0.474		0.664		-0.235		0.592		0.306		1.265	
		(0.65)		(0.961)		(0.973)		(1.436)		(2.561)		(1.29)	
	Intercept	36.160	**	34.129	**	37.827	**	36.327	**	38.626	**	27.596	**
		(1.542)		(2.825)		(4.218)		(3.038)		(3.181)		(3.381)	
	adj. R^2	0.669		0.664		0.766		0.694		0.466		0.716	
	n	7,186		2,245		1,527		1,178		1,308		928	
	N	202		63		43		33		37		26	

Note: ** p<.01; * p<.05. *Additional confounders:* average case duration (ACD), Months.

Specialties: General S: general surgery, Trauma S: trauma surgery, Uro S: urology, Ortho S: orthopedic surgery, Ear-Nose S: ear-nose-throat surgery.

Results of the second sensitivity analysis using the unbalanced sample supported the robustness of our findings. While the significance and direction of the decreasing effect remained stable, effect size slightly decreased (*Online Resource 2: Appendix C*). Also with regard to differences in the development across hospital types and specialties, results remain largely consistent with some decrease in effect size and significance.

For the second KPI considered, fixed effects analysis showed that average turnover time increased by 0.77 minutes in the first year of participation and by 1.92 minutes in the second year of participation, in comparison to the period before participation (p-value<0.01). This trend was especially consistent and statistically significant in maximum care hospitals, whose average turnover time increased by almost 2 minutes per year. A smaller but still statistically significant increase in turnover time was observed in main hospitals and general hospitals, albeit only for the second year of participation. University medical centers also presented positive coefficients, but with a much lower magnitude and with no statistical significance [Table 2]. All specialties included were affected by this trend except ear-nose-throat surgery, where turnover time development is not clear and not statistically significant. General surgery, trauma surgery and orthopedic surgery showed a statistically significant increase in turnover time by more than two minutes per change in the second year of participation.

Both sensitivity analyses confirmed these results. Results of the first sensitivity analysis using hospital specific fixed effects led to the same results as in the main specification (*Online Resource 2: Appendix B*). Results of the second sensitivity analysis using the unbalanced sample indicated that the increase of turnover time in main hospitals and maximum care hospitals appeared stronger and with a higher statistical significance (*Online Resource 2: Appendix C*). General surgery, trauma surgery and orthopedic surgery remained the three specialties mostly affected by the increase.

The results for raw utilization showed a slightly decreasing trend of raw utilization during the first year and an inversion of tendency starting from the second year of participation. The coefficients indicated however marginal and no statistically significant effects. Maximum care hospitals and main hospitals decreased their utilization at first, while increasing it during the second year, again slightly and not significantly. General hospitals increased their raw utilization regularly but not significantly from the beginning of participation, while university medical centers achieved a stable increase of almost 2 percent in the first year (p-value<0.05), which remained constant in the second year. No clear and statistically significant patterns emerged from the analysis for the different specialties included.

Results of both sensitivity analyses led to the same ambiguous results (*Online Resource 2: Appendix B/C*). Considering the unbalanced sample, the development of raw utilization in university medical centers appeared less strong and less significant.

4. Discussion

The analyses revealed a complex picture and no univocal trend in OR efficiency after entering a German benchmarking and reporting program for surgical process data. The effect depended upon the efficiency measure, the hospital type and the specialty considered.
For what concerns first-case tardiness, average delays of almost 11 minutes per case were observed in the sample, which were significantly reduced in the two years after program entrance. Similar results, albeit with lower and less significant coefficients, were evidenced in the sensitivity analyses. The resulting effect is not very large, but it still evidences a decreasing effect in time and a thus a learning effect. As this study analyzed changes per case in one year, by assuming a linear continuation of these increases/decreases, also small effects will probably acquire a significant importance in real practice. These findings support prior research demonstrating that OR management instruments, such as the regular use of KPIs, target setting and awareness raising strategies, foster higher time discipline of staff members and a lower first-case tardiness [9, 12, 23-26].
Turnover time showed a general statistically significant increase of almost one minute per year on average. This result was not only against any expectation, but also against the efforts of OR management in reducing turnover times. The possible reasons behind this result are diverse and need further attention. First of all, increases in turnover time might be caused by changes in unobserved variables, such as severity of patient illnesses and difficulty of surgeries, age of patients, decreasing staff levels and introduction of new documentation procedures and regulations [6, 10, 14, 27]. As a proxy for taking these factors into account, we included average case duration, which, however, might not fully capture all those changes. A possible reason for a lack of improvement in this key aspect of OR efficiency could be related to the fact that significantly shorter turnover times can be achieved only through structural interventions and additional staff to allow parallel processing, not only through behavioral changes and concerted efforts [2]. Given a widespread reluctance to change and the insecurity regarding cost-efficiency of these measures in the German research setting, probably they were only seldom introduced in the analyzed hospitals [2, 28-30].

Overall these results indicate a general absence of efficient solutions for the reduction of turnover times.

Raw utilization showed only a marginal and not significant increase in participant hospitals in the period considered. While this result might indicate a tendency to a marginal improvement of utilization in participating hospitals, it should be interpreted carefully. Despite the high relevance and the regular use of raw utilization as KPI in the practice, its intrinsic dynamics and its relevance for measuring and comparing OR efficiency are often questioned, especially for its lack of specificity in the inter-hospital comparison, as already highlighted by Faiz et al. [31-33].

A further step in the analysis focused on the development of the chosen KPIs in the different specialties and hospital types. For the subgroup-analyses in different specialties, no specific trend emerged, suggesting that type of specialty does not systematically impact on the efficacy of participation. The subgroup-analyses for the hospital types showed some interesting developments. University medical centers showed the greatest effect from participation regarding first-case tardiness and raw utilization and the lowest increase regarding turnover time. These results could be due to their major openness and engagement as pioneers of OR management and of KPI analysis, not only in the German context [1], but also in other countries [6,9], in comparison especially with other big non-academic hospitals (maximum care hospitals). These results suggest that motivation and hospital culture can increase the efficacy of participation and of analysis tools [8].

This study represents a first attempt at investigating perioperative efficiency trends in Germany in a large number of hospitals, especially after the entrance in a benchmarking and reporting program. While the analyses delivered first insights in this topic from a wider perspective, it also has some limitations which should be mentioned. First, the sample of hospitals included represents only a minority of German hospitals, which deliberately participated to the program and which could cause a serious positive selection bias. Furthermore, no data from a control group of non-participating hospitals were available, making the isolation of a participation effect difficult. Nonetheless, the database used to investigate the research question is the largest data pool available in Germany on this topic, based on routinely documented surgical process data in the anesthesia/nursing database of participating hospitals, posing no problems concerning a possible “Hawthorne-Effect” [23].

The analysis results furnish numerous starting points for further research. First, further investigations should take into account a joint analysis of the three KPIs considered and it should be complemented with further information, regarding volume of operations, accurate risk adjustment information, costs and revenues of departments. Second, further research methods should consider the implementation of dynamic panel data models in order to eliminate possible serial correlation effects in the results. Third, further research should focus on the unexpected increasing trend in turnover times. The impact of ageing and staff shortages should be analyzed, especially in order to unveil potential long-term consequences of this development on the overall OR efficiency and OR capacity availability.

This study represents a first attempt to investigate OR efficiency trends of German hospitals after entering a benchmarking and reporting program for surgical process data. The analysis helped to observe the efficiency trends over a period of three years, helping to detect which areas of OR efficiency are effectively covered by OR management strategies given the availability of regular and reliable controlling and comparison instruments. Results showed a complex picture and no clear and univocal trend, which depended on the aspect of OR efficiency, the hospital type and the specialty considered.

Compliance with ethical standards

Disclosure of potential conflicts of interest

Dr. Bialas is CEO of the external IT provider which gathers the data and administers the platform. He contributed with important intellectual content for the formulation of the research question and the interpretation of the results. He did not have any influence in the study design or analysis methodology.

Informed consent/IRB approval

Does not apply. All hospitals delivering data for the benchmarking program agreed to the use of anonymous data for scientific purposes. These were accessed by SP in the period between October 2015 and March 2016 in order to carry on the analysis.

Supplementary material

Online Resource 1: Key Performance Indicators

The supplementary online resource entails a detailed account of Key Performance Indicators used in the benchmarking program. (*Appendix A*)

Online Resource 2: Sensitivity Analysis Results Appendix

The supplementary online resource entails the full results of the two sensitivity analyses: the analysis of the main dataset with hospital specific fixed effects (*Appendix B*) and the analysis of the unbalanced dataset with all participants (*Appendix C*).

References

1. Baumgart A, Schupfer G, Welker A, Bender HJ, Schleppers A (2010) Status quo and current trends of operating room management in Germany. Current opinion in anaesthesiology 23 (2):193-200. doi:10.1097/ACO.0b013e328336b8b4
2. Marjamaa R, Vakkuri A, Kirvelä O (2008) Operating room management: why, how and by whom? Acta Anaesth Scand 52 (5):596-600
3. Bauer M, Wäschle RM, Rüggeberg J, Meyer HJ, Taube C, Diemer M, Schuster M (2016) The German Perioperative Procedural Time Glossary - A concerted recommendation of the German societies of anaesthesiology, surgery and operating room management. Anaesthesiol Intensivmed 57:669-683
4. Donham RT (1998) Defining measurable OR-PR scheduling, efficiency, and utilization data elements: the Association of Anesthesia Clinical Directors procedural times glossary. Int Anesthesiol Clin 36 (1):15-30
5. Marco AP, Hart S (2001) Cross-Industry Benchmarking: Is It Applicable to the Operating Room? Qual Manag Health Ca 9 (2):1-5
6. Gordon T, Paul S, Lyles A, Fountain J (1988) Surgical unit time utilization review: resource utilization and management implications. J Med Syst 12 (3):169-179
7. Malapero RJ, Gabriel RA, Gimlich R, Ehrenfeld JM, Philip BK, Bates DW, Urman RD (2015) An anesthesia medication cost scorecard–concepts for individualized feedback. J Med Syst 39 (5):48
8. Berry M, Berry-Stolzle T, Schleppers A (2008) Operating room management and operating room productivity: the case of Germany. Health Care Manag Sc 11 (3):228-239
9. van Veen-Berkx E, Elkhuizen SG, Kalkman CJ, Buhre WF, Kazemier G (2014) Successful interventions to reduce first-case tardiness in Dutch university medical centers: results of a nationwide operating room benchmark study. American journal of surgery 207 (6):949-959. doi:10.1016/j.amjsurg.2013.09.025
10. van Veen-Berkx E, Elkhuizen SG, van Logten S, Buhre WF, Kalkman CJ, Gooszen HG, Kazemier G (2015) Enhancement opportunities in operating room utilization; with a statistical appendix. The Journal of surgical research 194 (1):43-51.e41-42. doi:10.1016/j.jss.2014.10.044
11. Cardoen B, Demeulemeester E, Beliën J (2010) Operating room planning and scheduling: A literature review. Eur J Oper Res 201 (3):921-932
12. Chen Y, Gabriel RA, Kodali BS, Urman RD (2016) Effect of anesthesia staffing ratio on first-case surgical start time. J Med Syst 40 (5):1-6
13. Kodali BS, Kim D, Bleday R, Flanagan H, Urman RD (2014) Successful strategies for the reduction of operating room turnover times in a tertiary care academic medical center. The Journal of surgical research 187 (2):403-411
14. McDowell J, Wu A, Ehrenfeld JM, Urman RD (2017) Effect of the Implementation of a New Electronic Health Record System on Surgical Case Turnover Time. J Med Syst 41 (3):42
15. Porta CR, Foster A, Causey MW, Cordier P, Ozbirn R, Bolt S, Allison D, Rush R (2013) Operating room efficiency improvement after implementation of a postoperative team assessment. The Journal of surgical research 180 (1):15-20
16. Van Houdenhoven M, Hans EW, Klein J, Wullink G, Kazemier G (2007) A norm utilisation for scarce hospital resources: evidence from operating rooms in a Dutch university hospital. J Med Syst 31 (4):231-236
17. Schuster M, Bertheau S, Taube C, Bialas E, Bauer M (2014) Overlapping anaesthesia induction and perioperative turnover times An analysis of incidences and time expenditure in German hospitals based on 54,750 turnovers in 43 OR suites reporting to the benchmark

programme of the BDA/BDC and VOPM. ANASTHESIOLOGIE & INTENSIVMEDIZIN 55:654-661
18. Schuster M, Pezzella M, Taube C, Bialas E, Diemer M, Bauer M Verzögerungen beim morgendlichen Operationsbeginn.
19. Gabriel RA, Gimlich R, Ehrenfeld JM, Urman RD (2014) Operating room metrics score card—creating a prototype for individualized feedback. Journal of medical systems 38 (11):144
20. Gabriel RA, Wu A, Huang C-C, Dutton RP, Urman RD (2016) National incidences and predictors of inefficiencies in perioperative care. Journal of clinical anesthesia 31:238-246
21. Wooldridge JM (2010) Econometric analysis of cross section and panel data. MIT press,
22. Luedi MM, Kauf P, Mulks L, Wieferich K, Schiffer R, Doll D (2016) Implications of Patient Age and ASA Physical Status for Operating Room Management Decisions. Anesthesia and analgesia 122 (4):1169-1177. doi:10.1213/ane.0000000000001187
23. Overdyk FJ, Harvey SC, Fishman RL, Shippey F (1998) Successful strategies for improving operating room efficiency at academic institutions. Anesthesia and analgesia 86 (4):896-906
24. Panni M, Shah S, Chavarro C, Rawl M, Wojnarwsky P, Panni J (2013) Improving operating room first start efficiency–value of both checklist and a pre-operative facilitator. Acta Anaesth Scand 57 (9):1118-1123
25. Truong A, Tessler MJ, Kleiman SJ, Bensimon M (1996) Late operating room starts: experience with an education trial. Canadian journal of anaesthesia = Journal canadien d'anesthesie 43 (12):1233-1236. doi:10.1007/bf03013431
26. Wachtel RE, Dexter F (2009) Influence of the operating room schedule on tardiness from scheduled start times. Anesthesia and analgesia 108 (6):1889-1901
27. Schuster M, Kotjan T, Fiege M, Goetz A (2008) Influence of resident training on anaesthesia induction times. Brit J Anaesth 101 (5):640-647
28. Bender H, Waschke K, Schleppers A (2004) Are turnover times a measure of effective operating room management? Anaesthesiol Intensivmed 45 (9):529-535
29. Grote R, Sydow K, Radke J, Leuchtmann D, Menzel M, Walleneit A, Hudde T (2011) More operations per day through reducing turnover times: fact or fiction? Anaesthesiol Intensivmed 52
30. Grote R, Sydow K, Menzel M, Hunziker S, Schüpfer G (2009) Überlappende Einleitung. Der Anaesthesist 58 (10):1045-1047
31. Al-Benna S (2012) The impact of late-starts and overruns on theatre utilisation rates. J Perioper Pract 22 (6):197-199
32. Dexter F, Abouleish AE, Epstein RH, Whitten CW, Lubarsky DA (2003) Use of operating room information system data to predict the impact of reducing turnover times on staffing costs. Anesthesia and analgesia 97 (4):1119-1126
33. Faiz O, Tekkis P, McGuire A, Papagrigoriadis S, Rennie J, Leather A (2008) Is theatre utilization a valid performance indicator for NHS operating theatres? BMC health services research 8:28. doi:10.1186/1472-6963-8-28

Operating room efficiency before and after entrance in a benchmarking program for surgical process data

Sara Pedron, MSc

Institute for Health Economics and Health Care Management, Helmholtz Zentrum München, Germany. Ingolstädter Landstraße 1, 85764 Neuherberg (DE)

Vera Winter*, MA, PhD

Department of Political Science and Public Management, University of Southern Denmark, Denmark. Campusvej 55, 5230 Odense (DK)

Eva-Maria Oppel , MSc, PhD
Hamburg Center for Health Economics, ~~University of Hamburg~~Universität Hamburg, Germany. Esplanade 36, 20354 Hamburg (DE)

Enno Bialas, MD in Anesthesiology

digmed Datenmanagement im Gesundheitswesen GmbH. Flachsland 23, 22083 Hamburg (DE)

*corresponding author:

Vera Winter, MA, PhD

+45 6550 1884

winter@sam.sdu.dk

Field Code Changed

ABSTRACT

Background: Operating room (OR) efficiency continues to be a high priority for hospitals. In this context the concept of benchmarking has gained increasing importance as a means to improve OR performance. The aim of this study was to investigate whether and how participation in a benchmarking and reporting program for surgical process data was associated with a change in OR efficiency, measured through raw utilization, turnover times, and first-case tardiness.

Materials and Methods: The main analysis is based on panel data from 202 surgical departments in German hospitals, which were derived from the largest database for surgical process data in Germany. Panel regression modelling was applied.

Results: Results revealed no clear and univocal trend of participation in a benchmarking and reporting program for surgical process data. The largest trend was observed for first-case tardiness. In contrast to expectations, turnover times showed a generally increasing trend during participation. For raw utilization no clear and statistically significant trend could be evidenced. Subgroup analyses revealed differences in effects across different hospital types and department specialties.

Conclusions: Participation in a benchmarking and reporting program and thus the availability of reliable, timely and detailed analysis tools to support the OR management seemed to be correlated especially with an increase in the timeliness of staff members regarding first-case starts. The increasing trend in turnover time revealed the absence of effective strategies to improve this aspect of OR efficiency in German hospitals and could have meaningful consequences for the medium- and long-run capacity planning in the OR.

Keywords: Operating room, Utilization, Performance indicators, Benchmarking, Efficiency

1. Introduction

Health care providers are continuously striving to achieve higher levels of efficiency in response to health care challenges. Considerable attention within hospitals is dedicated to the operating room (OR) area, as one of the most important sources of hospital revenues and costs. Furthermore, OR process quality and safety is pivotal for the successful provision of hospital care [1,2]. As a result, increased attention in research and practice has been given to explore means to improve overall OR efficiency. In this context, the concept of benchmarking on standardized key performance indicators (KPIs) is increasingly applied to the health care industry in order to monitor OR efficiency and to improve performance [3-6]. Several benchmarking and reporting programs were developed in the last years in many European countries. The largest one in Germany, namely the *VOPM, BDA/DGAI, BDC benchmarking and reporting program*[1], was developed in 2010 and promoted by the main German professional associations of surgeons, anesthesiologists and OR managers. By entering the program, participants commit themselves to the monthly delivery of OR key process times and to a regular communication and update of structural hospital characteristics of surgical units (i.e. block time, allocated capacities, first case start target times), delivering also data from the year(s) previous to participation. The program is administered by an external IT provider, which ensures the compliance with current safety and privacy regulations and the plausibility of data. Participating hospitals have to pay an administrative participation fee per case to the IT provider.~~Participation is voluntary, against payment, and administered by an external IT provider, which ensures the compliance with current safety and privacy regulations and the plausibility of data~~. The pooled data are then analyzed through several established and standardized key performance indicators (KPIs), which are made available on a web-based dashboard [3]. The main objective of this benchmarking and reporting program is to allow a reliable, objective and regular performance evaluation of the own perioperative processes and the identification of differences in OR efficiency among different facilities~~competitors~~. As such participation itself has no direct effect on OR performance. However, external, reliable and detailed analysis tools could prove useful in

[1] Namely: VOPM (*Verband für OP-Management*, Association for OR management), BDA/DGAI (*Berufsverband Deutscher Anästhesisten*, Professional Association of German Anaesthesiologists & *Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin*, German Association of Anesthesiology and Intensive Care), BDC (*Berufsverbad der Deutschen Chirurgen*, Professional Association of German Surgeons).

targeting both critical aspects of low acceptance and poor use of analysis and controlling tools, in detecting inefficiencies, in increasing the accountability and involvement of staff members, all critical elements which are still a problem in many hospitals [1,6,7].

Prior research investigating the introduction and impact of KPIs as controlling and bench-marking tools focuses preponderantly on case studies at single institutions [1]. Thus, research exploring the impact of KPI benchmarking tools in a multiple institution setting and trends in OR efficiency in Germany remains scarce [8-10].

Addressing this research gap, the aim of this study is to investigate trends in OR efficiency of German hospitals participating in a benchmarking and reporting program. Specifically, we test whether and how hospitals differ in their OR efficiency before and after they enter a benchmarking program. In line with prior research and with the standard procedures issued by relevant stakeholders in the OR area, we define OR efficiency in terms of the following KPIs: raw utilization, turnover times, and first-case tardiness [3, 6, 10-16]. If OR efficiency improved after entering the program, this could indicate that participating in a benchmarking and reporting program helps to raise awareness and develop effective strategies to realize efficiency gains. Further, the results of our study enable detecting which areas might need further attention and development of the analysis tools. Finally, fruitful avenues for further research can be depicted.

2. Data and Methods

The study is a retrospective longitudinal evaluation of routinely documented data from the nursing/anesthesia database of each hospital, derived from the *VOPM*, *BDA/DGAI*, *BDC benchmarking and reporting program* for the period January 2009-July 2015. Previous studies have relied on this data already, corroborating its validity [17,18].

Data from the five most common surgical departments, namely general surgery (General S), trauma surgery (Trauma S), urology (Uro S), orthopedic surgery (Ortho S) and ear-nose-throat surgery (Ear-Nose S), were selected for the study as monthly average values for each department. Hospitals were then grouped into four hospital categories of similar size and complexity: general hospitals (General H, <5,000 cases/year), main hospitals (Main H, 5,000 -10,000 cases/year), maximum care hospitals (Max H, >10,000 cases/year), and university medical centers (Uni H, >10,000 cases/year and teaching/research activities). Furthermore, only departments from hospitals with more than 50 beds and departments which operated on average at least four times a month were included.

For the present study, observations from one year before entrance to two years after entrance in the program were selected, leading to a total of 36 monthly observations for each department. For periods which dated back more than one year prior to entrance, operational data were scarcely available and the communication of structural characteristics (capacity allocations and planning agreements) was not accurate. By the time the research was started, most hospitals took part in the program for no longer than two years. Therefore, we limited the participation period considered in this study to two years. The database used for the main analysis is thus a semi-balanced panel data, which included hospitals which participated at least two years in the program by July 2015 and whose process data for one year before entrance were fully available. The semi-balanced structure results due to some missing monthly observations and from outlier correction[2].

To measure OR efficiency we focus on three KPIs, which are widely used in theory and practice as important metrics of OR efficiency [3, 6, 10-16]. First-case tardiness is defined as the difference between the unit specific start target time and the incision time of the first case of the day. Turnover time is defined as the time between suture of one case and incision of the next patient, i.e. as suture-incision-time. Raw utilization is defined as total incision-suture time as a percentage of total allocated capacity. It should be stressed that there exist several definitions of these KPIs. For example, turnover times are sometimes defined as the time where the OR remains empty between two patients [19,20]. In this paper we use the definitions currently in use in Germany, which focuses on the time between suture of one case and incision of the next patient, i.e., suture-incision-time as turnover time [3]. For a more detailed definition please see *Appendix A*."

To estimate our panel data models, we performed weighted fixed effect regression with clustering at the department level and heteroscedasticity robust standard errors. Based on the results of a robust Hausman test, we decided to perform weighted fixed effect regression modelling rather than random effects modelling [21]. In order to account for differences across hospital type and specialty, separate regressions for each group were carried out.

The main analysis was carried out estimating fixed effects at department level, considering each department as an independent unit. To ensure the integrity and robustness of our results,

[2] Original observations at daily level which exceeded the lower and upper five percentiles were left out of monthly average calculations.

two sensitivity analyses were performed: First, to test whether results might be altered by unobserved variance at the hospital level, we re-ran our analyses using hospital fixed effects, with clustering at the hospital level (*Online Resource* 2~~1~~: ~~*Appendix A*~~*Appendix B*). Second, we re-ran our models using the full unbalanced dataset, including all hospitals participating in the program without the participation constraint of two years (unbalanced panel of 550 departments from 162 hospitals) (*Online Resource* 2~~1~~: ~~*Appendix B*~~*Appendix C*).
The basic specification represented a static model with department specific fixed effects:

$$KPI_{i,t} = \beta_0 + \delta_0\, Year(1)_t + \delta_1\, Year(2)_t + \beta_1 ACD_{i,t} + \sum_{m=2}^{12} \gamma_m (Month_{i,t})_m + \alpha_i + u_{i,t}$$

Where:
i=1, 2,…202
t=-11, -10…+24
m=1,2, …12

The coefficients on the year dummy variables, *Year(1)* and *Year(2)*, indicate the difference in the level of KPI with respect to the year before entrance [*Year(0)*] and hence provide insights into the development of efficiency in the years after entering the program.
Average case duration (*ACD*) and month dummies (*Month*) were included as control variables. Prior research showed that *ACD* is positively correlated with case complexity [12,22], making the former a good proxy variable for otherwise unavailable information for risk adjustment. Furthermore, in German hospitals, efficiency is lower during summer or during months with several public or school holidays due to lower staff availability and flexibility. All analyses were performed using the Stata 14 statistical software.

3. Results

Average first-case tardiness before entrance was 11.31 (±8.64) minutes per case, significantly decreasing in the first two years of participation, according to the paired t-test (Table 1). The decrease can be witnessed for almost all hospital types and specialties for both years, yet with a quite large variation in magnitude. Average turnover time before entrance was 56.64 (±14.01) minutes per change, showing a statistically significant increasing trend after entrance in the program. Again, there is little variation in the direction of this trend across different hospital types and specialties, but some variation in magnitude. Average raw utilization before entrance was 47.03 (±12.91) percent. Overall, no significant differences between mean values for the consecutive years could be found. For most sub-analyses on hospital type and on specialty level, changes are insignificant as well.

According to the regression analysis, the first two years of participation in the benchmarking program were overall associated with a progressive and statistically significant decrease in average first-case tardiness (Tables 2-3). This trend led on average to a 1.3 minutes reduction of delays per case during the second year of participation compared with the year before participation.

Table 1: descriptive statistics and paired t-test results for the different hospital types and specialties considered.

	Late starts (min)					**Turnover time (min)**					**Raw utilization (%)**				
	Y(0)	**Δ Y(1) - Y(0)**		**Δ Y(2) - Y(1)**		**Y(0)**	**Δ Y(1) - Y(0)**		**Δ Y(2) - Y(1)**		**Y(0)**	**Δ Y(1) - Y(0)**		**Δ Y(2) - Y(1)**	
	Mean (SD)	**Diff (SE)**		**Diff (SE)**		**Mean (SD)**	**Diff (SE)**		**Diff (SE)**		**Mean (SD)**	**Diff (SE)**		**Diff (SE)**	
Overall	11.31 (±8.64)	-0.83 (0.25)	**	-1.31 (0.25)	**	56.54 (±14.01)	0.95 (0.41)	*	2.14 (0.41)	**	47.03 (±12.91)	-0.13 (0.36)		0.37 (0.36)	
Hospital type															
General H	10.69 (±8.5)	-0.5 (0.42)		0 (0.45)		49.13 (±9.1)	0.33 (0.47)		1.37 (0.48)	**	43.53 (±12.75)	0.2 (0.64)		0.69 (0.65)	
Main H	11.81 (±8.38)	-1.75 (0.46)	**	-2.29 (0.46)	**	53.23 (±10.69)	0.99 (0.58)		2.18 (0.6)	**	48.83 (±11.7)	-0.67 (0.62)		0.15 (0.62)	
Max H	9.12 (±7.28)	0.87 (0.51)		-0.68 (0.52)		57.52 (±9.34)	1.49 (0.64)	*	3.53 (0.68)	**	48.41 (±15.52)	-2.39 (0.91)	**	-1.37 (1.05)	
Uni H	12.81 (±9.47)	-1.39 (0.57)	*	-1.96 (0.56)	**	71.19 (±16.12)	0.03 (0.97)		0.12 (0.99)		48.58 (±11.43)	1.67 (0.69)	*	1.48 (0.65)	*
Specialty															
General S	11.26 (±7.79)	-0.93 (0.38)	*	-1.18 (0.39)	**	55.91 (±13.95)	1 (0.71)		2.36 (0.72)	**	50.94 (±12.42)	-0.42 (0.59)		0.61 (0.6)	
Trauma S	13.28 (±9.21)	-0.59 (0.56)		0.26 (0.56)		59.2 (±14.28)	1.31 (0.88)		3.33 (0.9)	**	47.02 (±13.12)	-0.51 (0.79)		-0.41 (0.81)	
Uro S	13.02 (±8.1)	-1.52 (0.55)	**	-2.3 (0.57)	**	59.92 (±12.94)	0.53 (0.87)		1.35 (0.88)		45.1 (±14.54)	-0.54 (0.9)		0.25 (0.97)	
Ortho S	9.91 (±8.5)	-0.27 (0.65)		-2.09 (0.62)	**	58.41 (±11.67)	1.65 (0.86)		2.45 (0.85)	**	45.45 (±11.1)	0.86 (0.83)		0.48 (0.8)	
Ear-Nose S	7.39 (±9.28)	-0.83 (0.75)		-2.29 (0.75)	**	48.17 (±13.79)	-0.05 (1.14)		0.36 (1.12)		42.33 (±10.76)	0.45 (0.86)		1.05 (0.85)	

Note: ** p<.01; * p<.05.

Hospital types: General H: general hospitals, Main H: main hospitals, Max H: maximum care hospitals, Uni H: university medical center.

Specialties: General S: general surgery, Trauma S: trauma surgery, Uro S: urology, Ortho S: orthopedic surgery, Ear-Nose S: ear-nose-throat surgery.

Similarly, a statistically significant and progressively decreasing trend ~~were~~ was identified in both university medical centers (Uni H) and main hospitals (Main H). In these groups, benchmarking participation was associated with an average decrease of delays of more than 2 minutes per case during the second year. The trends in maximum care hospitals (Max H) and general hospitals (General H) were not statistically significant.

Table 2: fixed effects regression analysis results, overall and for the different hospital types.

Outcome		OVERALL		General H		Main H		Max H		Uni H	
Late starts (min)	**Year (1)**	-0.837	**	-0.395		-1.714	*	0.989		-1.587	**
		(0.32)		(0.377)		(0.701)		(0.837)		(0.552)	
	Year (2)	-1.276	**	0.046		-2.286	**	-0.143		-2.221	**
		(0.389)		(0.443)		(0.838)		(1.119)		(0.652)	
	Intercept	8.209	**	5.970	**	7.858	**	6.381	**	10.652	**
		(0.923)		(1.204)		(2.183)		(2.172)		(1.484)	
	adj. R^2	0.7532		0.8046		0.7486		0.5363		0.8143	
	N	7,252		2,268		2,052		1,240		1,692	
	N	202		63		57		35		47	
Turnover time (min)	**Year (1)**	0.774	**	0.471		0.844		1.841	**	0.158	
		(0.262)		(0.351)		(0.544)		(0.454)		(0.68)	
	Year (2)	1.925	**	1.528	**	1.902	*	3.855	**	0.864	
		(0.398)		(0.468)		(0.76)		(0.854)		(1.088)	
	Intercept	50.025	**	42.732	**	49.116	**	49.593	**	62.229	**
		(1.109)		(1.136)		(1.621)		(1.34)		(2.239)	
	adj. R^2	0.9013		0.8483		0.8455		0.8256		0.8775	
	N	7,272		2,268		2,052		1,260		1,692	
	N	202		63		57		35		47	
Raw utilization (%)	**Year (1)**	-0.040		0.175		-0.487		-2.612		1.964	*
		(0.472)		(0.579)		(0.787)		(1.748)		(0.856)	
	Year (2)	0.474		0.478		0.232		-1.562		1.957	
		(0.65)		(0.964)		(0.967)		(2.443)		(1.193)	
	Intercept	36.160	**	28.403	**	33.410	**	38.263	**	40.369	**
		(1.542)		(3.091)		(3.399)		(4.885)		(2.201)	
	adj. R^2	0.6688		0.7275		0.708		0.5293		0.647	
	n	7,186		2,250		2,024		1,238		1,674	
	N	202		63		57		35		47	

Note: ** p<.01; * p<.05. *Additional confounders:* average case duration (ACD), Months.

Hospital types: General H: general hospitals, Main H: main hospitals, Max H: maximum care hospitals, Uni H: university medical center.

Four of the five specialties included showed a generally decreasing trend in first-case tardiness in the first two years of participation. However, this decreasing trend was statistically significant only for three of the specialties (Ortho S, Uro S and Ear-Nose S) starting from the second year.

Results of the first sensitivity analysis using hospital specific fixed effects further supported the robustness of our findings. As shown in ~~Appendix A~~Appendix B, results remain stable in terms of direction, significance, and effect size (*Online Resource 2: ~~Appendix A~~Appendix B*).

Table 3: fixed effects regression analysis results, overall and for the different specialties included.

Outcome		OVERALL		General S		Trauma S		Ortho S		Uro S		Ear-Nose S	
Late starts (min)	Year (1)	-0.837	**	-0.942		-0.518		-0.199		-1.657		-0.907	
		(0.32)		(0.513)		(0.641)		(0.847)		(1.013)		(0.811)	
	Year (2)	-1.276	**	-1.271		0.665		-2.104	*	-2.625	*	-2.076	*
		(0.389)		(0.663)		(0.716)		(0.997)		(1.133)		(0.839)	
	Intercept	8.209	**	6.587	**	8.407	**	4.673	**	13.149	**	4.733	**
		(0.923)		(1.602)		(1.52)		(1.582)		(1.699)		(1.497)	
	adj. R^2	0.753		0.691		0.838		0.760		0.654		0.754	
	n	7,252		2,268		1,548		1,188		1,312		936	
	N	202		63		43		33		37		26	
Turnover time (min)	Y(1)	0.774	**	0.809		1.027		1.498	**	0.517		-0.359	
		(0.262)		(0.436)		(0.588)		(0.436)		(0.792)		(0.707)	
	Y(2)	1.925	**	2.123	**	2.956	**	2.072	**	1.362		0.214	
		(0.398)		(0.686)		(0.798)		(0.732)		(1.457)		(0.851)	
	Intercept	50.025	**	49.581	**	48.470	**	49.330	**	56.827	**	42.931	**
		(1.109)		(1.313)		(2.756)		(2.349)		(2.387)		(1.304)	
	adj. R^2	0.901		0.901		0.912		0.899		0.788		0.926	
	n	7,272		2,268		1,548		1,188		1,332		936	
	N	202		63		43		33		37		26	
Raw utilization (%)	Y(1)	-0.040		-0.159		-0.308		0.759		-0.490		0.253	
		(0.472)		(0.708)		(0.765)		(1.305)		(1.678)		(0.954)	
	Y(2)	0.474		0.664		-0.235		0.592		0.306		1.265	
		(0.65)		(0.961)		(0.973)		(1.436)		(2.561)		(1.29)	
	Intercept	36.160	**	34.129	**	37.827	**	36.327	**	38.626	**	27.596	**
		(1.542)		(2.825)		(4.218)		(3.038)		(3.181)		(3.381)	
	adj. R^2	0.669		0.664		0.766		0.694		0.466		0.716	
	n	7,186		2,245		1,527		1,178		1,308		928	
	N	202		63		43		33		37		26	

Note: ** p<.01; * p<.05. *Additional confounders:* average case duration (ACD), Months.

Specialties: General S: general surgery, Trauma S: trauma surgery, Uro S: urology, Ortho S: orthopedic surgery, Ear-Nose S: ear-nose-throat surgery.

Results of the second sensitivity analysis using the unbalanced sample supported the robustness of our findings. While the significance and direction of the decreasing effect remained stable, effect size slightly decreased (*Online Resource 2: ~~Appendix B~~Appendix C*). Also with regard to differences in the development across hospital types and specialties, results remain largely consistent with some decrease in effect size and significance.

For the second KPI considered, fixed effects analysis showed that average turnover time increased by 0.77 minutes in the first year of participation and by 1.92 minutes in the second year of participation, in comparison to the period before participation (p-value<0.01). This trend was especially consistent and statistically significant in maximum care hospitals, whose average turnover time increased by almost 2 minutes per year. A smaller but still statistically significant increase in turnover time was observed in main hospitals and general hospitals, albeit only for the second year of participation. University medical centers also presented positive coefficients, but with a much lower magnitude and with no statistical significance [Table 2]. All specialties included were affected by this trend except ear-nose-throat surgery, where turnover time development is not clear and not statistically significant. General surgery, trauma surgery and orthopedic surgery showed a statistically significant increase in turnover time by more than two minutes per change in the second year of participation.

Both sensitivity analyses confirmed these results. Results of the first sensitivity analysis using hospital specific fixed effects led to the same results as in the main specification (*Online Resource 2~~1~~: ~~Appendix A~~Appendix B*). Results of the second sensitivity analysis using the unbalanced sample indicated that the increase of turnover time in main hospitals and maximum care hospitals appeared stronger and with a higher statistical significance (*Online Resource 2~~1~~: ~~Appendix B~~Appendix C*). General surgery, trauma surgery and orthopedic surgery remained the three specialties mostly affected by the increase.

The results for raw utilization showed a slightly decreasing trend of raw utilization during the first year and an inversion of tendency starting from the second year of participation. The coefficients indicated however marginal and no statistically significant effects. Maximum care hospitals and main hospitals decreased their utilization at first, while increasing it during the second year, again slightly and not significantly. General hospitals increased their raw utilization regularly but not significantly from the beginning of participation, while university medical centers achieved a stable increase of almost 2 percent in the first year (p-value<0.05), which remained constant in the second year. No clear and statistically significant patterns emerged from the analysis for the different specialties included.

Results of both sensitivity analyses led to the same ambiguous results (*Online Resource 2: ~~Appendix A~~Appendix B/C~~B~~*). Considering the unbalanced sample, the development of raw utilization in university medical centers appeared less strong and less significant.

4. Discussion

The analyses revealed a complex picture and no univocal trend in OR efficiency after entering a German benchmarking and reporting program for surgical process data. The effect depended upon the efficiency measure, the hospital type and the specialty considered.

For what concerns first-case tardiness, average delays of almost 11 minutes per case were observed in the sample, which were significantly reduced in the two years after program entrance. Similar results, albeit with lower and less significant coefficients, were evidenced in the sensitivity analyses. The resulting effect is not very large, but it still evidences a decreasing effect in time and a thus a learning effect. As this study analyzed changes per case in one year, by assuming a linear continuation of these increases/decreases, also small effects will probably acquire a significant importance in real practice. These findings support prior research demonstrating that OR management instruments, such as the regular use of KPIs, target setting and awareness raising strategies, foster higher time discipline of staff members and a lower first-case tardiness [9, 12, 23-26].

Turnover time showed a general statistically significant increase of almost one minute per year on average. This result was not only against any expectation, but also against the efforts of OR management in reducing turnover times. The possible reasons behind this result are diverse and need further attention. First of all, increases in turnover time might be caused by changes in unobserved variables, such as severity of patient illnesses and difficulty of surgeries, age of patients, decreasing staff levels and introduction of new documentation procedures and regulations [6, 10, 14, 27]. As a proxy for taking these factors into account, we included average case duration, which, however, might not fully capture all those changes. A possible reason for a lack of improvement in this key aspect of OR efficiency could be related to the fact that significantly shorter turnover times can be achieved only through structural interventions and additional staff to allow parallel processing, not only through behavioral changes and concerted efforts [2]. Given a widespread reluctance to change and the insecurity regarding cost-efficiency of these measures in the German research setting, probably they were only seldom introduced in the analyzed hospitals [2, 28-30].

Overall these results indicate a general absence of efficient solutions for the reduction of turnover times.

Raw utilization showed only a marginal and not significant increase in participant hospitals in the period considered. While this result might indicate a tendency to a marginal improvement of utilization in participating hospitals, it should be interpreted carefully. Despite the high relevance and the regular use of raw utilization as KPI in the practice, its intrinsic dynamics and its relevance for measuring and comparing OR efficiency are often questioned, especially for its lack of specificity in the inter-hospital comparison, as already highlighted by Faiz et al. [31-33].

A further step in the analysis focused on the development of the chosen KPIs in the different specialties and hospital types. For the subgroup-analyses in different specialties, no specific trend emerged, suggesting that type of specialty does not systematically impact on the efficacy of participation. The subgroup-analyses for the hospital types showed some interesting developments. University medical centers showed the greatest effect from participation regarding first-case tardiness and raw utilization and the lowest increase regarding turnover time. These results could be due to their major openness and engagement as pioneers of OR management and of KPI analysis, not only in the German context [1], but also in other countries [6,9], in comparison especially with other big non-academic hospitals (maximum care hospitals). These results suggest that motivation and hospital culture can increase the efficacy of participation and of analysis tools [8].

This study represents a first attempt at investigating ~~OR~~ perioperative efficiency trends in Germany in a large number of hospitals, especially after the entrance in a benchmarking and reporting program. While the analyses delivered first insights in this topic from a wider perspective, it also has some limitations which should be mentioned. First, the sample of hospitals included represents only a minority of German hospitals, which deliberately participated to the program and which could cause a serious positive selection bias. Furthermore, no data from a control group of non-participating hospitals were available, making the isolation of a participation effect difficult. Nonetheless, the database used to investigate the research question is the largest data pool available in Germany on this topic, based on routinely documented surgical process data in the anesthesia/nursing database of participating hospitals, posing no problems concerning a possible "Hawthorne-Effect" [23].

The analysis results furnish numerous starting points for further research. First, further investigations should take into account a joint analysis of the three KPIs considered and it should be complemented with further information, regarding volume of operations, accurate risk adjustment information, costs and revenues of departments. Second, further research methods should consider the implementation of dynamic panel data models in order to eliminate possible serial correlation effects in the results. Third, further research should focus on the unexpected increasing trend in turnover times. The impact of ageing and staff shortages should be analyzed, especially in order to unveil potential long-term consequences of this development on the overall OR efficiency and OR capacity availability.

This study represents a first attempt to investigate OR efficiency trends of German hospitals after entering a benchmarking and reporting program for surgical process data. The analysis helped to observe the efficiency trends over a period of three years, helping to detect which areas of OR efficiency are effectively covered by OR management strategies given the availability of regular and reliable controlling and comparison instruments. Results showed a complex picture and no clear and univocal trend, which depended on the aspect of OR efficiency, the hospital type and the specialty considered.

Compliance with ethical standards

Disclosure of potential conflicts of interest

Dr. Bialas is CEO of the external IT provider which gathers the data and administers the platform. He contributed with important intellectual content for the formulation of the research question and the interpretation of the results. He did not have any influence in the study design or analysis methodology.

Informed consent/IRB approval

Does not apply. All hospitals delivering data for the benchmarking program agreed to the use of anonymous data for scientific purposes. These were accessed by SP in the period between October 2015 and March 2016 in order to carry on the analysis.

Supplementary material

Online Resource 1: Key Performance Indicators

The supplementary online resource entails a detailed account of Key Performance Indicators used in the benchmarking program. (*Appendix A*)

Online Resource 21: Sensitivity Analysis Results Appendix

The supplementary online resource entails the full results of the two sensitivity analyses: the analysis of the main dataset with hospital specific fixed effects (~~*Appendix*~~ *Appendix B*~~*A*~~) and the analysis of the unbalanced dataset with all participants (*Appendix* *C*~~*B*~~).

References

1. Baumgart A, Schupfer G, Welker A, Bender HJ, Schleppers A (2010) Status quo and current trends of operating room management in Germany. Current opinion in anaesthesiology 23 (2):193-200. doi:10.1097/ACO.0b013e328336b8b4
2. Marjamaa R, Vakkuri A, Kirvelä O (2008) Operating room management: why, how and by whom? Acta Anaesth Scand 52 (5):596-600
3. Bauer M, Wäschle RM, Rüggeberg J, Meyer HJ, Taube C, Diemer M, Schuster M (2016) The German Perioperative Procedural Time Glossary - A concerted recommendation of the German societies of anaesthesiology, surgery and operating room management. Anaesthesiol Intensivmed 57:669-683
4. Donham RT (1998) Defining measurable OR-PR scheduling, efficiency, and utilization data elements: the Association of Anesthesia Clinical Directors procedural times glossary. Int Anesthesiol Clin 36 (1):15-30
5. Marco AP, Hart S (2001) Cross-Industry Benchmarking: Is It Applicable to the Operating Room? Qual Manag Health Ca 9 (2):1-5
6. Gordon T, Paul S, Lyles A, Fountain J (1988) Surgical unit time utilization review: resource utilization and management implications. J Med Syst 12 (3):169-179
7. Malapero RJ, Gabriel RA, Gimlich R, Ehrenfeld JM, Philip BK, Bates DW, Urman RD (2015) An anesthesia medication cost scorecard–concepts for individualized feedback. J Med Syst 39 (5):48
8. Berry M, Berry-Stolzle T, Schleppers A (2008) Operating room management and operating room productivity: the case of Germany. Health Care Manag Sc 11 (3):228-239
9. van Veen-Berkx E, Elkhuizen SG, Kalkman CJ, Buhre WF, Kazemier G (2014) Successful interventions to reduce first-case tardiness in Dutch university medical centers: results of a nationwide operating room benchmark study. American journal of surgery 207 (6):949-959. doi:10.1016/j.amjsurg.2013.09.025
10. van Veen-Berkx E, Elkhuizen SG, van Logten S, Buhre WF, Kalkman CJ, Gooszen HG, Kazemier G (2015) Enhancement opportunities in operating room utilization; with a statistical appendix. The Journal of surgical research 194 (1):43-51.e41-42. doi:10.1016/j.jss.2014.10.044
11. Cardoen B, Demeulemeester E, Beliën J (2010) Operating room planning and scheduling: A literature review. Eur J Oper Res 201 (3):921-932
12. Chen Y, Gabriel RA, Kodali BS, Urman RD (2016) Effect of anesthesia staffing ratio on first-case surgical start time. J Med Syst 40 (5):1-6
13. Kodali BS, Kim D, Bleday R, Flanagan H, Urman RD (2014) Successful strategies for the reduction of operating room turnover times in a tertiary care academic medical center. The Journal of surgical research 187 (2):403-411
14. McDowell J, Wu A, Ehrenfeld JM, Urman RD (2017) Effect of the Implementation of a New Electronic Health Record System on Surgical Case Turnover Time. J Med Syst 41 (3):42
15. Porta CR, Foster A, Causey MW, Cordier P, Ozbirn R, Bolt S, Allison D, Rush R (2013) Operating room efficiency improvement after implementation of a postoperative team assessment. The Journal of surgical research 180 (1):15-20
16. Van Houdenhoven M, Hans EW, Klein J, Wullink G, Kazemier G (2007) A norm utilisation for scarce hospital resources: evidence from operating rooms in a Dutch university hospital. J Med Syst 31 (4):231-236
17. Schuster M, Bertheau S, Taube C, Bialas E, Bauer M (2014) Overlapping anaesthesia induction and perioperative turnover times An analysis of incidences and time expenditure in German hospitals based on 54,750 turnovers in 43 OR suites reporting to the benchmark

programme of the BDA/BDC and VOPM. ANASTHESIOLOGIE & INTENSIVMEDIZIN 55:654-661
18. Schuster M, Pezzella M, Taube C, Bialas E, Diemer M, Bauer M Verzögerungen beim morgendlichen Operationsbeginn.
19. Gabriel RA, Gimlich R, Ehrenfeld JM, Urman RD (2014) Operating room metrics score card—creating a prototype for individualized feedback. Journal of medical systems 38 (11):144
20. Gabriel RA, Wu A, Huang C-C, Dutton RP, Urman RD (2016) National incidences and predictors of inefficiencies in perioperative care. Journal of clinical anesthesia 31:238-246
21. Wooldridge JM (2010) Econometric analysis of cross section and panel data. MIT press,
22. Luedi MM, Kauf P, Mulks L, Wieferich K, Schiffer R, Doll D (2016) Implications of Patient Age and ASA Physical Status for Operating Room Management Decisions. Anesthesia and analgesia 122 (4):1169-1177. doi:10.1213/ane.0000000000001187
23. Overdyk FJ, Harvey SC, Fishman RL, Shippey F (1998) Successful strategies for improving operating room efficiency at academic institutions. Anesthesia and analgesia 86 (4):896-906
24. Panni M, Shah S, Chavarro C, Rawl M, Wojnarwsky P, Panni J (2013) Improving operating room first start efficiency–value of both checklist and a pre-operative facilitator. Acta Anaesth Scand 57 (9):1118-1123
25. Truong A, Tessler MJ, Kleiman SJ, Bensimon M (1996) Late operating room starts: experience with an education trial. Canadian journal of anaesthesia = Journal canadien d'anesthesie 43 (12):1233-1236. doi:10.1007/bf03013431
26. Wachtel RE, Dexter F (2009) Influence of the operating room schedule on tardiness from scheduled start times. Anesthesia and analgesia 108 (6):1889-1901
27. Schuster M, Kotjan T, Fiege M, Goetz A (2008) Influence of resident training on anaesthesia induction times. Brit J Anaesth 101 (5):640-647
28. Bender H, Waschke K, Schleppers A (2004) Are turnover times a measure of effective operating room management? Anaesthesiol Intensivmed 45 (9):529-535
29. Grote R, Sydow K, Radke J, Leuchtmann D, Menzel M, Walleneit A, Hudde T (2011) More operations per day through reducing turnover times: fact or fiction? Anaesthesiol Intensivmed 52
30. Grote R, Sydow K, Menzel M, Hunziker S, Schüpfer G (2009) Überlappende Einleitung. Der Anaesthesist 58 (10):1045-1047
31. Al-Benna S (2012) The impact of late-starts and overruns on theatre utilisation rates. J Perioper Pract 22 (6):197-199
32. Dexter F, Abouleish AE, Epstein RH, Whitten CW, Lubarsky DA (2003) Use of operating room information system data to predict the impact of reducing turnover times on staffing costs. Anesthesia and analgesia 97 (4):1119-1126
33. Faiz O, Tekkis P, McGuire A, Papagrigoriadis S, Rennie J, Leather A (2008) Is theatre utilization a valid performance indicator for NHS operating theatres? BMC health services research 8:28. doi:10.1186/1472-6963-8-28

Operating room efficiency before and after entrance in a benchmarking program for surgical process data

Sara Pedron, Vera Winter, Eva-Maria Oppel, Enno Bialas*

Journal of Medical Systems

Online Resource 1/Appendix A: Key Performance Indicators

Figure A.1: representation of relevant time points and periods of the perioperative process as defined in the *German Perioperative Procedural Time Glossary [3]*.

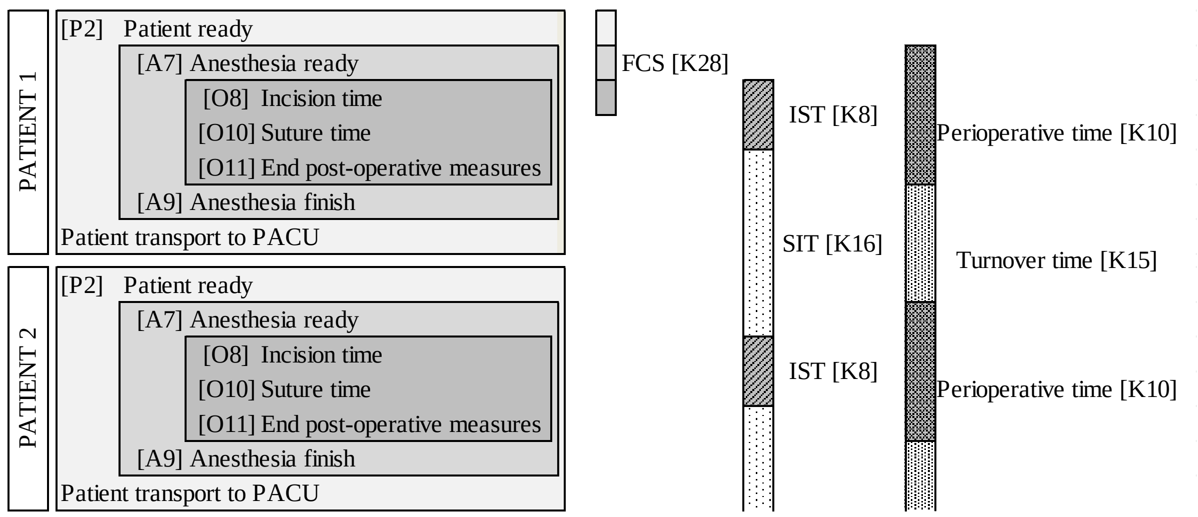

Table A.1: definition of the most common key performance indicators (KPIs) as defined in the *German Perioperative Procedural Time Glossary [3]*. Grey shaded lines report the definitions used in the present study.

KPI	Code*	Formula	Name	Description
First case start times	**K28a**	$P2_A$-$P2_P$	on time starts (Patient ready)	Minutes between planned patient ready ($P2_P$) and actual patient ready time ($P2_A$)
	K28b	$A7_A$-$A7_P$	on time starts (Anesthesia ready)	Minutes between planned anesthesia ready ($A7_P$) and actual anesthesia ready ($A7_A$)
	K28c	$O8_A$-$O8_P$	*on time starts (Incision time)*	*Minutes between planned incision time ($O8_P$) and actual incision time ($O8_A$)*
Turnover times	***K16***	$O8_2$ - $O10_1$	*suture-incision time (SIT)*	*Minutes between suture of one patient ($O10_1$) and incision of the next patient*
	K15b	$A7_2$ - $O11_1$	turnover time anesthesia	Minutes between end of post-operative measures on one patient ($O11_1$) and anesthesia ready of the next patient ($A7_2$)
Operative times	**K8**	O10 - O8	incision-suture time (IST)	Minutes between incision of one patient (O8) and suture of the same patient (O10)
	K10	O11 - A7	perioperative time	Minutes between anesthesia ready of one patient (A7) and end of post-operative measures of the same patient (O11)
Planned capacities	**K18**		planned capacity/block time	Total time allocated for a surgery unit in one day
Utilization	***K20***	*Sum(K8)/K18*	*Raw utilization*	*Sum of incision-suture time (K8) as a share of total planned capacity (K18)*
		Sum(K10)/K18	Adjusted utilization	Sum of perioperative time (K10) as a share of total planned capacity (K18)

Notes: A= actual; P=planned; 1=patient one; 2=patient two(following patient);
* code number as indicated in the *German Perioperative procedural Time Glossary [3]*.

Operating room efficiency before and after entrance in a benchmarking program for surgical process data

Sara Pedron, Vera Winter, Eva-Maria Oppel, Enno Bialas*

Journal of Medical Systems

Online Resource 2: Sensitivity Analysis Results Appendix

*corresponding author:

Vera Winter, MA, PhD

Department of Political Science and Public Management, University of Southern Denmark, Denmark. Campusvej 55, 5230 Odense (DK)

+45 6550 1884

winter@sam.sdu.dk

Appendix B: Sensitivity Analysis with hospital specific fixed effects

Table B.1: fixed effects regression with hospital specific fixed effects, divided for hospital type.

Outcome		Overall		General H		Main H		Max H		Uni H	
Late starts (min)	**Year (1)**	-0.814	*	-0.379		-1.655	*	0.925		-1.518	**
		(0.32)		(0.38)		(0.697)		(0.85)		(0.546)	
	Year (2)	-1.273	**	0.033		-2.285	**	-0.212		-2.130	**
		(0.397)		(0.442)		(0.859)		(1.12)		(0.676)	
	Intercept	6.679	**	5.368	**	2.012		1.789		10.542	**
		(1.261)		(1.725)		(1.891)		(2.18)		(2.515)	
	adj. R^2	0.468		0.597		0.411		0.349		0.497	
	N	7252		2268		2052		1240		1692	
	N	64		25		17		12		10	
Turnover time (min)	**Year (1)**	0.758	**	0.402		1.036		1.620	**	0.216	
		(0.265)		(0.367)		(0.569)		(0.48)		(0.68)	
	Year (2)	1.828	**	1.369	**	2.033	*	3.542	**	0.649	
		(0.402)		(0.483)		(0.797)		(0.836)		(1.091)	
	Intercept	45.453	**	35.618	**	37.355	**	51.574	**	59.263	**
		(2.281)		(3.009)		(4.312)		(4.963)		(5.246)	
	adj. R^2	0.629		0.593		0.530		0.504		0.363	
	N	7272		2268		2052		1260		1692	
	N	64		25		17		12		10	
Raw utilization (%)	**Year (1)**	-0.095		0.178		-0.638		-2.613		1.846	*
		(0.467)		(0.58)		(0.809)		(1.738)		(0.83)	
	Year (2)	0.395		0.485		0.112		-1.559		1.739	
		(0.644)		(0.951)		(0.981)		(2.464)		(1.135)	
	Intercept	37.118	**	30.100	**	32.121	**	41.035	**	42.791	**
		(1.593)		(2.649)		(2.715)		(3.461)		(1.947)	
	adj. R^2	0.387		0.476		0.431		0.281		0.344	
	n	7186		2250		2024		1238		1674	
	N	64		25		17		12		10	

Note: ** p<.01; * p<.05. *Additional confounders:* average case duration (ACD), Months.

Hospital types: General H: general hospitals, Main H: main hospitals, Max H: maximum care hospitals, Uni H: university medical centers.

Table B.2: fixed effects regression with hospital specific fixed effects, divided for specialty.

Outcome		Overall		General S		Trauma S		Ortho S		Uro S		Eear-Nose S	
Late starts (min)	**Year (1)**	-0.814	*	-0.926		-0.519		-0.112		-1.566		-0.907	
		(0.32)		(0.512)		(0.641)		(0.856)		(1.005)		(0.811)	
	Year (2)	-1.273	**	-1.270		0.669		-2.126	*	-2.397	*	-2.076	*
		(0.397)		(0.665)		(0.715)		(0.998)		(1.183)		(0.839)	
	Intercept	6.679	**	5.822	**	8.432	**	1.140		10.177	**	4.733	**
		(1.261)		(0.865)		(1.518)		(2.251)		(1.487)		(1.497)	
	adj. R^2	0.4679		0.6800		0.8368		0.7534		0.5819		0.7539	
	N	7,252		2,268		1,548		1,188		1,312		936	
	N	64		59		42		30		33		26	
Turnover time (min)	**Year (1)**	0.758	**	0.786		1.024		1.486	**	0.194		-0.359	
		(0.265)		(0.44)		(0.587)		(0.437)		(0.777)		(0.707)	
	Year (2)	1.828	**	2.123	**	2.971	**	2.079	**	0.946		0.214	
		(0.402)		(0.682)		(0.8)		(0.729)		(1.433)		(0.851)	
	Intercept	45.453	**	51.467	**	48.598	**	49.943	**	58.097	**	42.931	**
		(2.281)		(2.026)		(2.768)		(1.574)		(2.901)		(1.304)	
	adj. R^2	0.6292		0.8902		0.906		0.8984		0.728		0.926	
	n	7,272		2,268		1,548		1,188		1,332		936	
	N	64		59		42		30		33		26	
Raw utilization (min)	**Year (1)**	-0.095		-0.386		-0.311		0.783		-0.432		0.253	
		(0.467)		(0.709)		(0.766)		(1.313)		(1.698)		(0.954)	
	Year (2)	0.395		0.571		-0.230		0.544		0.301		1.265	
		(0.644)		(0.99)		(0.974)		(1.424)		(2.579)		(1.29)	
	Intercept	37.118	**	45.225	**	37.459	**	33.647	**	36.508	**	27.596	**
		(1.593)		(1.954)		(4.347)		(1.853)		(3.256)		(3.381)	
	adj. R^2	0.3868		0.6409		0.756		0.6807		0.4067		0.716	
	n	7,186		2,245		1,527		1,178		1,308		928	
	N	64		59		42		30		33		26	

Note: ** p<.01; * p<.05. *Additional confounders:* average case duration (ACD), Months.

Specialties: General S: general surgery, Trauma S: trauma surgery, Uro S: urology, Ortho S: orthopedic surgery, Ear-Nose S: ear-nose-throat surgery.

Appendix C: Sensitivity analysis with unbalanced panel data

Table C.1: descriptive statistics of the full unbalanced dataset.

	Late starts (min)			Turnover time (min)			Raw utilization (%)		
	Y(0)	**Δ Y(1) - Y(0)**	**Δ Y(2) - Y(1)**	**Y(0)**	**Δ Y(1) - Y(0)**	**Δ Y(2) - Y(1)**	**Y(0)**	**Δ Y(1) - Y(0)**	**Δ Y(2) - Y(1)**
	Mean (SD)	**Diff (SE)**	**Diff (SE)**	**Mean (SD)**	**Diff (SE)**	**Diff (SE)**	**Mean (SD)**	**Diff (SE)**	**Diff (SE)**
Overall	11.31 (±8.64)	-0.83 (0.25) **	-1.31 (0.25) **	56.54 (±14.01)	0.95 (0.41) *	2.14 (0.41) **	47.03 (±12.91)	-0.13 (0.36)	0.37 (0.36)
Level of care									
General H	10.69 (±8.5)	-0.5 (0.42)	0 (0.45)	49.13 (±9.1)	0.33 (0.47)	1.37 (0.48) **	43.53 (±12.75)	0.2 (0.64)	0.69 (0.65)
Main H	11.81 (±8.38)	-1.75 (0.46) **	-2.29 (0.46) **	53.23 (±10.69)	0.99 (0.58)	2.18 (0.6) **	48.83 (±11.7)	-0.67 (0.62)	0.15 (0.62)
Max H	9.12 (±7.28)	0.87 (0.51)	-0.68 (0.52)	57.52 (±9.34)	1.49 (0.64) *	3.53 (0.68) **	48.41 (±15.52)	-2.39 (0.91) **	-1.37 (1.05)
Uni H	12.81 (±9.47)	-1.39 (0.57) *	-1.96 (0.56) **	71.19 (±16.12)	0.03 (0.97)	0.12 (0.99)	48.58 (±11.43)	1.67 (0.69) *	1.48 (0.65) *
Specialty									
General S	11.26 (±7.79)	-0.93 (0.38) *	-1.18 (0.39) **	55.91 (±13.95)	1 (0.71)	2.36 (0.72) **	50.94 (±12.42)	-0.42 (0.59)	0.61 (0.6)
Trauma S	13.28 (±9.21)	-0.59 (0.56)	0.26 (0.56)	59.2 (±14.28)	1.31 (0.88)	3.33 (0.9) **	47.02 (±13.12)	-0.51 (0.79)	-0.41 (0.81)
Uro S	13.02 (±8.1)	-1.52 (0.55) **	-2.3 (0.57) **	59.92 (±12.94)	0.53 (0.87)	1.35 (0.88)	45.1 (±14.54)	-0.54 (0.9)	0.25 (0.97)
Ortho S	9.91 (±8.5)	-0.27 (0.65)	-2.09 (0.62) **	58.41 (±11.67)	1.65 (0.86)	2.45 (0.85) **	45.45 (±11.1)	0.86 (0.83)	0.48 (0.8)
Ear-Nose S	7.39 (±9.28)	-0.83 (0.75)	-2.29 (0.75) **	48.17 (±13.79)	-0.05 (1.14)	0.36 (1.12)	42.33 (±10.76)	0.45 (0.86)	1.05 (0.85)

Note: ** p<.01; * p<.05.
Hospital types: General H: general hospitals, Main H: main hospitals, Max H: maximum care hospitals, Uni H: university medical centers.
Specialties: General S: general surgery, Trauma S: trauma surgery, Uro S: urology, Ortho S: orthopedic surgery, Ear-Nose S: ear-nose-throat surgery.

Table C.2: fixed effects regression with unbalanced panel, divided for hospital type.

Outcome		Overall		General H		Main H		Max H		Uni H	
Late starts (min)	**Year (1)**	-0.571	*	-0.222		-1.162	**	0.062		-0.516	
		(0.226)		(0.292)		(0.371)		(0.576)		(0.498)	
	Year (2)	-0.934	**	-0.161		-1.119		-0.553		-1.542	*
		(0.318)		(0.443)		(0.585)		(0.865)		(0.636)	
	Intercept	7.801	**	6.063	**	6.409	**	7.000	**	8.834	**
		(0.84)		(0.941)		(1.083)		(1.363)		(1.597)	
	adj. R^2	0.7800		0.7690		0.7470		0.7588		0.8224	
	n	14,409		4,128		5,007		2,182		3,092	
	N	549		153		200		83		113	
Turnover time (min)	**Year (1)**	0.834	**	0.647		0.908	**	1.475	**	0.493	
		(0.182)		(0.339)		(0.297)		(0.314)		(0.45)	
	Year (2)	1.821	**	1.348	**	2.234	**	3.318	**	0.765	
		(0.293)		(0.446)		(0.506)		(0.66)		(0.739)	
	Intercept	49.669	**	42.402	**	47.315	**	49.110	**	62.072	**
		(0.728)		(1.417)		(1.218)		(1.114)		(1.368)	
	adj. R^2	0.9030		0.8457		0.8449		0.8786		0.8852	
	n	14,422		4,135		5,018		2,174		3,095	
	N	549		154		200		82		113	
Raw utilization (%)	**Year (1)**	0.027		0.291		-0.401		-0.992		0.975	
		(0.302)		(0.485)		(0.503)		(1.041)		(0.567)	
	Year (2)	0.530		1.024		-0.114		-0.751		1.654	
		(0.478)		(0.808)		(0.741)		(1.713)		(0.906)	
	Intercept	35.067	**	31.944	**	29.412	**	33.689	**	40.725	**
		(1.364)		(2.276)		(2.306)		(3.716)		(2.257)	
	adj. R^2	0.7080		0.7324		0.7343		0.5899		0.7243	
	n	14,256		4,111		4,963		2,147		3,035	
	N	548		153		200		82		113	

Note: ** p<.01; * p<.05. *Additional confounders:* average case duration (ACD), Months.

Hospital types: General H: general hospitals, Main H: main hospitals, Max H: maximum care hospitals, Uni H: university medical centers.

Table C.3: fixed effects regression with unbalanced panel, divided for specialty.

Outcome		Overall		General S		Trauma S		Ortho S		Uro S		Ear-Nose S	
Late starts (min)	Year (1)	-0.571	*	-0.804		-0.067		-0.602		-0.797		-0.553	
		(0.226)		(0.429)		(0.401)		(0.613)		(0.6)		(0.495)	
	Year (2)	-0.934	**	-1.372	*	0.714		-1.658	*	-1.492		-0.973	
		(0.318)		(0.617)		(0.51)		(0.724)		(0.787)		(0.823)	
	Intercept	7.801	**	8.764	**	8.490	**	5.932	**	9.749	**	2.173	
		(0.84)		(2.136)		(0.867)		(1.016)		(1.408)		(1.249)	
	adj. R^2	0.7801		0.6953		0.8510		0.7381		0.7672		0.7563	
	n	14,409		4,310		3,297		2,104		2,607		2,091	
	N	549		167		127		77		98		80	
Turnover time (min)	Year (1)	0.834	**	0.695	*	1.104	**	1.185	**	0.754		0.415	
		(0.182)		(0.341)		(0.392)		(0.338)		(0.49)		(0.404)	
	Year (2)	1.821	**	1.910	**	2.484	**	2.470	**	1.395		0.469	
		(0.293)		(0.515)		(0.605)		(0.581)		(1.066)		(0.543)	
	Intercept	49.669	**	49.289	**	50.906	**	51.612	**	53.800	**	42.585	**
		(0.728)		(1.032)		(1.83)		(1.727)		(1.642)		(0.907)	
	adj. R^2	0.9028		0.8968		0.8947		0.8995		0.8265		0.9253	
	n	14,422		4,314		3,298		2,117		2,599		2,094	
	N	549		168		127		77		97		80	
Raw utilization (%)	Year (1)	0.027		-0.307		-0.052		0.351		0.148		0.308	
		(0.302)		(0.409)		(0.504)		(1.087)		(1.006)		(0.596)	
	Year (2)	0.530		0.470		0.200		0.346		1.009		0.815	
		(0.478)		(0.742)		(0.763)		(1.248)		(1.765)		(0.924)	
	Intercept	35.067	**	33.464	**	31.364	**	39.274	**	35.950	**	32.772	**
		(1.364)		(2.018)		(2.951)		(2.82)		(2.412)		(5.687)	
	adj. R^2	0.7084		0.7192		0.7510		0.6667		0.5827		0.7696	
	n	14,256		4,271		3,224		2,101		2,581		2,079	
	N	548		167		126		77		98		80	

Note: ** p<.01; * p<.05. *Additional confounders:* average case duration (ACD), Months.

Specialties: General S: general surgery, Trauma S: trauma surgery, Uro S: urology, Ortho S: orthopedic surgery, Ear-Nose S: ear-nose-throat surgery.

Dear Editor,

First of all, we would like to express our gratitude to you and the reviewers for their very helpful comments. By adopting the suggestions and responding to the comments, we believe that we have substantially improved the scientific quality of the enclosed manuscript.
On the following pages, we detail our responses to each specific comment and refer to the related changes in the manuscript. To ensure traceability about changes in the manuscript we use the track changes option in the manuscript. Further, besides a marked copy we provide a clean copy of the manuscript.

Thank you again for your time and effort you gave this paper.

Sincerely,
Vera Winter (corresponding author)

COMMENTS FOR THE AUTHOR and RESPONSES

Reviewer #1:

In this retrospective review of a large database, the authors investigate the impact of participation in an OR metrics benchmarking database on OR efficiency metrics. Overall this is a well performed study with a robust statistical analysis. I have a few questions/comments for the authors.

Response: Thank you for your detailed and constructive comments and suggestions. We have addressed each in turn below. We appreciate the time and attention you gave to this paper.

1. Please define what VOPM, BDA/DGAI, BDC stand for

Response: VOPM stands for “Verband für OP-Management” (“Association for OR management”), BDA stands for “Berufsverband Deutscher Anästhesisten” (“Professional Association of German Anaesthesiologists”), DGAI stands for “Deutsche Gesellschaft für Anästhesiologie und Intensivmedizin” (“German Association of Anesthesiology and Intensive Care”), BDC stands for “Berufsverbad der Deutschen Chirurgen” (“Professional Association of German Surgeons”). These associations represent all staff groups involved in the OR and are the main associations in Germany. Due to space limitations, we cannot include these definitions in the text. However, we provide detailed explanations of these acronyms in footnote 1.

2. Introduction: what do you mean by, "against payment"?

Response: The benchmarking and reporting program is administered by an external IT provider. While participation is voluntary for the hospitals, there is a participation fee (administrative fee per case). We used the terminology “against payment” to inform the reader about the existence of this participation fee. We regret any uncertainty on this terminology. In revising the Introduction, we now state: “The program is administered by an external IT provider, which ensures the compliance with current safety and privacy regulations and the plausibility of data. Participating hospitals have to pay an administrative participation fee per case to the IT provider.” (page 3)

3. Introduction: KPIs - can you include examples of these metrics in a table? or as an appendix? or refer to a website where they're all listed and their definitions?

Response: We included a table in the appendix, where selected KPIs evaluated in the Benchmarking program are listed and explained (Online Supplementary Material 1/Appendix A: Key Performance Indicators).

We also included references to original publications (in German and English):

3. Bauer M, Wäschle RM, Rüggeberg J, Meyer HJ, Taube C, Diemer M, Schuster M (2016) The German Perioperative Procedural Time Glossary - A concerted recommendation of the German societies of anaesthesiology, surgery and operating room management. Anaesthesiol Intensivmed 57:669-683

4. Donham RT (1998) Defining measurable OR-PR scheduling, efficiency, and utilization data elements: the Association of Anesthesia Clinical Directors procedural times glossary. Int Anesthesiol Clin 36 (1):15-30

4. Intro: instead of "competitors' use "different facilities" or something similar.

Response: Thank you for this suggestion. We now use different facilities instead of competitors (page 3).

5. Intro: please list your study Aims towards the end of the introduction, not in the middle

Response: In response to your comment, we now include the study's aims and contributions as separate paragraph at the end of the introduction.

6. Data and methods: do you have an IRB approval for this research? I know it doesn't require patient consent, but still may need an approval.

Response: The analyzed data were transferred from the hospitals to the IT provider in anonymous form. In the terms and conditions to participate in the program all participants agreed to the use of data in aggregated form for scientific purposes. The authors evaluated only data at monthly level. In Germany, no IRB approval is needed for this kind of studies given the consent of each participating institution. We highlighted this point in the final section "informed consent/IRB approval".

7. Methods: can you describe how each hospital reports the data - is it manually entered, automatically transferred via EHR? How clean/accurate is it- have other studies been published looking at the integrity of the data and the database? Can you describe the infrastructure of the database which houses all collected data?

Response: For documentation purposes, each hospital in Germany must collect several information regarding surgery times in the hospital nursing/anesthesia database. They are normally recorded manually in a standardized electronic form by nurses/anesthesiologists during the whole permanence of the patient in the OR. Data regarding the allocation of capacities and the start times in the morning, which are regularly updated by hospitals, are also transferred regularly to the IT provider, which updates its database consecutively.

No study has already been published regarding the integrity of these data. However its quality is guaranteed by:

(1) Regular plausibility checks regarding the time sequence of surgery points are carried out automatically by the IT provider every time that new data comes in. Each and every error is reported to the hospitals which have to provide explanation or a correct version of the data.

(2) The data are then uploaded on an online evaluation tool, where actual data are compared with capacities and time goals. These data are used by hospitals for monthly checks of their performance: any not plausible or striking result is usually reported by the firm to the hospital. Usually changes in capacities/time goals/room closures are responsible for those errors, and are consequently updated by the provider.

The infrastructure of the database which houses all collected data is extremely complex and is managed by four IT specialists on a daily basis. Every month new data are delivered. For each case the information regarding each case is connected with the information regarding capacities and start time goals. Following standardized procedures, the matched data are then analyzed to produce relevant time periods or ratios (such as the distance between actual start time and planned start time of the first surgery, or the ratio of cases which started on time out of all first cases performed). They are then uploaded on the online platform, where participating hospitals can access and evaluate several KPIs. Further information is available upon request.

Furthermore, in the Data and Methods section we included this paragraph:

"Previous studies have relied on this data already, corroborating its validity [17,18]."

8. Methods: "average at least 4 times" - please explain. Do you mean hospitals doing at least 4 surgeries per month? Seems like a really low threshold. Also, can you comment on the percent of hospital type (by surgical volume) out of the total study sample you used?

Response: We excluded departments, not hospitals, carrying out less than four surgeries per month on average (i.e. operating less than one time a week on average). In our opinion this is a low but reasonable threshold. We changed the Methods section on order to make the distinction between hospital and department clear.

In the sensitivity analysis we used all available data on general surgery, trauma surgery, urology, orthopedic surgery and ear-nose-throat surgery departments (except hospitals which had less than 50 beds, departments which operated less than four times a month on average, which represented isolated exceptions). We included 552 departments from general surgery, trauma surgery, urology, orthopedic surgery and ear-nose-throat surgery departments

(~2,066,006 cases). For the main analysis we included all departments which participated regularly for two years and which delivered data for at least one year before participation. 202 of the 552 departments included in the sensitivity analysis satisfied this condition (37%) but they included more than 52% of the total observations (~1,078,085) in a balanced panel structure.

9. Methods: not entirely clear how you define on-time start - it is OK if your database uses a different definition, you just have to be clear and also can you comment whether ALL hospitals in Germany which participated use the same efficiency metrics?? On time start: Some use the time to induction complete, others incision made, still others getting in the OR by certain time in the morning.

Response: Thank you for this suggestion. A detailed definition of the types of KPIs that can be computed for the on time first case starts of the day was included in the Appendix as suggested in comment 3. As you already mentioned, there are several time points which can be considered for the on time start, namely: “patient ready”, “anesthesia ready” and “incision”.

In order to participate in the program, hospitals have to deliver at least two time points: incision and suture times of every case. This means that each and every participating hospital can evaluate its timeliness of first case starts regarding the incision point in comparison to all other participating institutions. The delivery of the other two time points is voluntary, but several hospitals choose to deliver them anyway, so that the majority is also able to analyze their timeliness with regard to the time points “patient ready” and “anesthesia ready” in comparison to the other hospitals. Generally, all metrics included in Appendix A are evaluated by participating hospitals in the benchmarking program, based on standardized data collection, delivery and computation.

In the Methods section, the KPI “first case start time” is defined as “the difference between the unit specific start target time and the incision time of the first case of the day”. We included a link to the *Appendix A* table in text in order to make the definition clearer.

10. Methods: turnover time (TOT): most definitions in the US (using AACD official definitions), at least, define TOT as the time when the OR is empty (last patient left the OR to new patient entering the OR). See and cite the following:

a. Gabriel RA, Wu A, Huang CC, Dutton RP, Urman RD. National incidences and predictors of inefficiencies in perioperative care. J Clin Anesth. 2016 Jun;31:238-46. PMID: 27185719
b. Gabriel RA, Gimlich R, Ehrenfeld JM, Urman RD. Operating room metrics score card-creating a prototype for individualized feedback. J Med Syst. 2014 Nov;38(11):144. PMID: 25315824

Response: Thank you for this comment and suggestion. We included this paragraph in the methods section:

"It should be stressed that there exist several definitions of these KPIs. For example, turnover times are sometimes defined as the time where the OR remains empty between two patients [19, 20]. In this paper we use the definitions currently in use in Germany, which focuses on the time between suture of one case and incision of the next patient, i.e., suture-incision-time as turnover time [3]. For a more detailed definition please see *Appendix A*."

11. Methods: weighted fixed effect regression modelling rather than random effects modelling. Please explain why.

Response: As a result of an insignificant Hausman test, both estimators yield similar results and can be used without any difference.

12. Results: Generally, is each hospital able to see the metrics from other hospitals?
Response: Yes, but in aggregated form on the online platform. What they usually see is the average/median, the benchmark and themselves. Generally, they can only see the value of the score of the benchmark for each KPI, they cannot see who the benchmark is. In this way they can position themselves in comparison all other institutions and to the best performer. If the benchmark hospitals agree, in some cases the IT provider can reveal who the benchmark is: in this way hospitals can form a cooperation and share solutions on similar problems which could enhance their efficiency.

13. page 7, line 48: should we was, not were

Response: Thanks for detecting this typo. We revised the sentence accordingly.

14. Discussion: "the resulting effect is not very large" - it is important to emphasize that small statistically significant differences may not have any importance in real practice

(improving TOTs by 1-2 mins cannot allow for another case to be scheduled, for example).

Response: Thank you for your comment. We agree that a 1-2 minute-improve may not have great importance in real practice to start with, yet we believe that it probably has relevance on a more long-term and aggregate basis. An increase of TOTs by one minute per case per year with on average 5 turnovers per day per department results in 5 minutes per day on average. A five minute saving might indeed allow performing another surgery in some occasions. Additionally, if the trend continues this way, this means that in five years each turnover will last 5 minutes longer: the same day with on average 5 changes per department will last in five years 25 minutes longer. Whether another case can be scheduled or not depends on how well the hospital is organized and which case we are dealing with but the magnitude of the effect should not be underestimated. We therefore included the paragraph in the discussion session:

"The resulting effect is not very large, but it still evidences a decreasing effect in time and a thus a learning effect. As this study analyzed changes per case in one year, by assuming a linear continuation of these increases/decreases, also small effects will probably acquire a significant importance in real practice."

15. page 12, line 24: first attempt in investigate in Germany? for US, see #10 (a) reference above.

Response: Thank you for this suggestion. We changed the paragraph as follows:

"This study represents a first attempt at investigating perioperative efficiency trends in Germany in a large number of hospitals, especially after the entrance in a benchmarking and reporting program."

The reference you suggested focused on other measures of postoperative efficacy. What we would like to stress in the above cited phrase is that we are the first ones in Germany to look at the *evolution* of *perioperative efficiency*, in hospitals which participated in a program. Several other studies focused on similar issues but we are not aware of any study looking at trends.

16. page 12, line 31: what is the actual percentages of hospitals that are included in this database?

Response: We included departments from 164 hospitals in the sensitivity analysis and from 64 hospitals in the main analysis. Out of the total 1956 hospitals in Germany in 2015[1], these represent 8% and 3% respectively. Out of all University Medical Centers in Germany (n=35[1]), the included UMCs (n=10) represented 29% of these institutions.

[1] Statistisches Bundesamt, 2016. *Gesundheit. Grunddaten der Krankenhäuser*. https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Krankenhaeuser/GrunddatenKrankenhaeuser2120611157004.pdf?__blob=publicationFile. Last accessed: 26.07.2017.