EUR J PHYS REHABIL MED 2015;51:803-13

Clinical outcome and cost effectiveness of inpatient rehabilitation after total hip and knee arthroplasty. A multi-centre cohort benchmarking study between nine rehabilitation departments in Rhineland-Palatinate (Western Germany)

C. BAULIG [1, 2], M. GRAMS [1], B. RÖHRIG [2], S. LINCK-ELEFTHERIADIS [2], F. KRUMMENAUER [1]

[1]Institute of Medical Biometry and Epidemiology Faculty of Health, Witten/Herdecke University Witten, Germany
[2]Medical Service of the Health Care Insurance Rhineland-Palatinate (MDK Rheinland-Pfalz), Alzey, Germany

Background. **Rehabilitation after hip and knee replacement surgery is effective, but cost intensive. To ensure consistent cost-effectiveness across different providers prospective cohort studies on its clinical and economic outcome are recommended. Comparisons alongside suitable outcome quality indicators enable to contrast different providers and — in case of clinically or economically relevant differences — reveal constructive approaches to quality improvement.**
Aim. **Therefore an external benchmarking for post-hospital curative treatment after hip and knee arthroplasty between nine inpatient rehabilitation departments in the German Federal Land Rhineland-Palatinate was implemented based on data acquired between 01/2007 and 12/2009.**
Design. **Multi-centre retrospective cohort benchmarking study.**
Setting. **Inpatient rehabilitation in nine rehabilitation departments**
Population. **A total of 8672 patients after unilateral hip arthroplasty (THA) and 8180 patients after unilateral knee replacement (TKA) surgery were investigated. The median age of the patients after hip arthroplasty varied between 71 and 75 years, after knee arthroplasty between 72 and 75 years. The department-wise proportions of female patients ranged from 62% to 77% (THA) and from 70% to 81% (TKA).**
Methods. **Data on clinical outcome of inpatient rehabilitation after THA and TKA were documented using the EVAReha® software. As primary indicator of clinical outcome quality the intra-individual pre-post change in the total Staffelstein Index was determined [%], as primary economic indicator the cost-normalized effect estimate (CNEE) was estimated [%/€] relating the individual Staffelstein Index change to a patient's respective direct costs [€] paid by the statutory health insurance.**
Results. **In the nine departments the median Staffelstein index increase after THA ranged between 18% to 31% corresponding to median CNEEs ranging from 9% to 15% Staffelstein increase per €1000 investment of the statutory health (Kruskal/Wallis P<0.001). After TKA the median Staffelstein index increase ranged between 19% to 34% corresponding to median CNEEs ranging from 9% to 16% Staffelstein increase per 1,000 € investment of the statutory health (Kruskal / Wallis P<0.001).**
Conclusion. **Inpatient follow-up treatment after hip and after knee replacement surgery demonstrated a statistically significant and both clinically and economically relevant heterogeneity between the departments. Additional exploratory analyses by means of adjustment with further predictors could not resolve these heterogeneities.**
Clinical Rehabilitation Impact. **Benchmark information on cost effectiveness would enable departments to learn from the "best" and to implement corresponding peer cooperation, to then identify own shortcomings and potentials, and thereby to analyze and optimize local processes.**

KEY WORDS: Rehabilitation - Treatment Outcome - Costs and Cost Analysis - Benchmarking.

Corresponding author: C. Baulig, Institute of Medical Biometry and Epidemiology, Faculty of Health, Witten/Herdecke University, Alfred Herrhausen Strasse 50, D- 58448 Witten, Germany.
E-mail: christine.baulig@uni-wh.de

Osteoarthritis is the most common joint disease [1] and, in Germany, concerns about two thirds of the 60-year-old population and over 80% of the 70-year-old country population.[2] In 2009 German clinics performed approximately 210,000 total hip arthroplasties (THA) and 175,000 total knee arthroplasties (TKA)[3] – Kurtz *et al.* [4] calculated for the year 2010 a number of 253,000 THA and 663,000 TKA in the USA. Thereby orthopedic endoprothetic surgery is one of the most common surgical procedures [5] and leads to notable annual economic investment for every health care system in both the initial surgery and in the postoperative care.

The posthospital curative treatment (PHCT) is a special form of medical rehabilitation which is affiliated to an inpatient hospital treatment.[6] In Germany, PHCT after orthopedic endoprothetics can alternatively be performed in an outpatient or inpatient rehabilitation department.[7] In 2009, a total number of 670,270 patients were treated in preventive and rehabilitation orthopedic departments - which were well ahead of the departments of internal medicine with nearly 500,000 patients.[8] The focus of rehabilitation goals is the "self-determination and participation in the life of society" based on the International Classification of Functioning, Disability and Health (ICF) of the World Health Organization (WHO). After endoprosthesis (TEP) surgical implantation, PHCT should begin no later than 14 days after discharge from the acute hospital. For the elderly, PHCT are usually reimbursed by health care insurers, where a major part of German patients refers to statutory health care insurances.

The increasing need for and the resulting increasing economic relevance of PHCT after arthroplasty, imply the need to ensure consistent cost effectiveness over various providers. As a consequence prospective benchmarking investigations on the clinical and economic outcome of the inpatient procedures' results are necessary and recommended. Benchmarking is a promising "instrument of competitive analysis" in health care research: limitations of resources more than ever leads to a competitive situation,[9] which made benchmarking methods become more important for the German health care system during the past decade.

In the present study a market-based benchmarking is carried out. The main characteristic of market-based benchmarking is the affiliation of the individual competitors to the same division using nearly homogeneous workflows implying the advantage of comparability. For inpatient PHCT after THA and TKA an external benchmarking study between 9 competing inpatient rehabilitation departments in the German Federal Land country of Rhineland-Palatinate was implemented in terms of a retrospective cohort study.

Materials and methods

Patient data

This retrospective multicenter cohort benchmarking study was conducted at 9 rehabilitation departments in Rhineland-Palatinate (Western Germany). Exclusively degenerative arthritis cases with a PHCT following THA or TKA were included. Data from patients who had to be re-transported to an acute hospital during rehabilitation due to medical complications were excluded from analysis. From 1/2007 to 12/2009 a total sample of 16,852 patients — 8672 after THA and 8180 after TKA — were recorded in these 9 rehabilitation clinics. For each patient the responsible department clinicians documented the "Staffelstein Score" at the start and the end of the inpatient rehabilitation and entered the raw score data into the "EVA-REHA®" software database (see below). From this database the data was sent to the Medical Service of the health care insurance Rhineland-Palatinate (MDK Rheinland-Pfalz) and stored in a data archive. The data was extracted and then provided anonymously in an output format of the statistical software SPSS® (version 18). With regard to the anonymous data collection, no vote of the local independent Ethics Committee was required; nevertheless, the local data protection officer of the Medical Service was involved, who gave a positive vote on the above study design by 09.14.2011. Furthermore, the Rhineland-Palatinate Ministry of Social and Family Affairs was informed about the project intention and gave a correspondingly positive vote by 09.14.2011. Any informed consent from human subjects was obtained as required.

The EVA-REHA® software

In 2004, the Medical Service of the Health Care Insurance Rhineland-Palatinate implemented a Windows Access® based remote data entry facility to

enable rehabilitation departments all over Germany for patient outcome data on the clinical situation of patients undergoing indoor PHCT in Rhineland-Palatinate. During 2007–2010 documentation by means of this "EVAReha®" software [10] was declared mandatory for any inpatient rehabilitation department in Rhineland-Palatinate under contract of statutory health care insurers. As a consequence, the resulting database can — at least for the statutory health care insurers' patients — be considered as an unselected complete census on clinical outcome quality in inpatient PHCT after orthopedic endoprothetic surgery in Rhineland-Palatinate.

Staffelstein Score

For follow-up documentation of objective clinical outcome and subjective patient-related outcome during rehabilitation after THA or TKA, Middeldorf and Casser proposed the "Staffelstein Score",[11] which is directly related to the Harris Hip Score usually applied in THA surveillance. For the three dimensions pain, ADL (activities of daily living) and joint function a series of 13 items cumulates a maximum of 120 points for best possible joint health (40 points for each dimension). For the following this 120 points maximum score was linearly transformed into a percentage range (minimum 0% and maximum +100%) and then referred to as a "Staffelstein Index". Note, that this transformation only reproduced the range of a utility Index but, regarding the functional perspective of the Staffelstein items, did not provide one in the formal sense. Nevertheless, the range of 100% seemed to be more appropriate, when relating the individual change in post to prerehabilitation Staffelstein individual assessments to the underlying financial investments.

Direct costs

Individual direct costs were estimated by multiplying the daily cost rate of the respective rehabilitation department with the individual residence time (days) for each patient in the respective clinic as documented in the EVAReha® database. The daily rates differ between individual rehabilitation facilities and result from annual negotiations between the health insurance and the respective rehabilitation department. For this investigation one had to refer to daily cost rates dating back to the year 2007, since no more recent rates were made available by the insurers. Nevertheless, the order of these rates was assured to not have changed relevantly between 2007 and 2009; in particular, the relative order of rates between departments could be assumed stable.

Clinical endpoint

The intra-individual pre-post change of the Staffelstein Index [%] (calculated as difference of both values) was the primary clinical endpoint of this study. It ranges from -100% to +100%, where a negative change ≤0% implies treatment failure due to post-rehabilitation worsening.

Economic endpoint

The cost-normalized effect estimate (CNEE) [%/€1000] was defined as the primary economic endpoint of this study. It was calculated as the quotient of the intra-individual Staffelstein Index change [%] to the respective patient's direct cost profile [€] as invested by the statutory health insurance. It reflects individual cost effectiveness of an individual patient's health care improvement in relation to the costs for this improvement: If, for example, a patient's Staffelstein Index increased from 25% to 75% at the end of inpatient treatment, and the health care insurer in charge invested a total of € 000 in this treatment, this patient's individual CNEE was then calculated 50%/€2000=25% Index increase per €1000 investment. A negative CNEE implies worsening of the Staffelstein Index despite substantial financial investment for the corresponding patient's post-operative indoor rehabilitation.

Statistical analysis

Evaluations were performed using the IBM® SPSS® 19. The clinical outcome and cost effectiveness data were analyzed separately for the THA and TKA patients. Data description was based on medians and quartiles, graphical representation on box plots.

For continuous outcome variables, differences between the 9 departments were tested by means of the Kruskal/Wallis global Test at a 5% significance level. As a consequence, the confirmatory analysis of this benchmark investigation, the global department comparison alongside the CNEE, was therefore based on a 5% Kruskal/Wallis Test. Pair-wise comparisons were then based on two sample Wilcoxon

Test; the results of these pair-wise Wilcoxon Test were considered exploratory and the resulting p-values formally adjusted for multiplicity.

To account for a possible association between the CNEE and cofactors such as age and gender patterns among the 9 rehabilitation departments, the respective department comparisons were furthermore stratified for these cofactors.

Results

For a total of 8672 THA-patients and 8180 TKA-patients treatment outcome data were documented by means of the EVAReha® software. Between the 9 departments, the median age of the patients after hip arthroplasty varied between 71 and 75 years, after knee arthroplasty between 72 and 75 years; the department-wise proportions of female patients ranged from 62% to 77% (THA) and from 70% to 81% (TKA). Table I demonstrates the socio-demographic characteristics of the overall THA and TKA cohorts.

THA

In the THA cohort a median Staffelstein increase of 24% (medians ranging from 18% to 31% between the departments) from a median index of 53% (medians ranging 42-61%) when entering indoor treatment to a median index of 78% (medians ranging 70-93%) at withdrawal was reported. The intra-individual changes differed locally significantly at the 5% level between the departments (Kruskal/Wallis Test P<0.001). In each department, the underlying patients showed notable variation in the primary clinical endpoint: the greatest index change in median was 31% (interquartile range for the respective department's patients 23-38%), the smallest change in median was 18% (patients' interquartile range 12-26%). Table II summarizes the distribution charac-

TABLE I.—*Sociodemographic data.*

Department code (u=urban area r=rural area)	THA				TKA			
	N.	Age [years		Gender	N	Age [years]		Gender
		Median	Min-Max	Female [%]		Median	Min-Max	Female [%]
1 (r)	704	73	32-93	69.7%	713	73	50-90	75.0%
2 (u)	459	73	36-89	66.9%	438	74	41-90	80.8%
3 (r)	521	74	43-95	65.6%	564	73	43-90	76.1%
7 (r)	361	71	31-87	62.0%	354	72	52-89	70.9%
8 (r)	994	72	28-93	67.3%	796	73	37-91	69.7%
11 (u)	3814	72	20-95	69.6%	3466	72	41-92	74.2%
12 (r)	921	73	23-96	63.7%	1005	73	43-90	72.7%
13 (r)	572	74	45-92	73.8%	504	74	48-107	79.6%
14 (u)	326	75	49-91	77.3%	340	75	44-89	77.4%
Total	8672	73	20-96	68.6%	8180	73	37-107	74.5%

TABLE II.—*THA – clinical outcome.*

Department code (u=urban area, r=rural area)	N	Pre-SSI [%]		Post-SSI [%]		Change in SSI [%]	
		Median	Q1-Q3	Median	Q1-Q3	Median	Q1-Q3
1 (r)	704	56	48-64	79	70-86	22	17-28
2 (u)	459	42	37-48	70	63-78	27	20-34
3 (r)	521	61	53-70	93	85-94	26	20-37
7 (r)	361	51	43-57	81	74-90	31	23-38
8 (r)	994	50	42-56	75	68-79	25	18-31
11 (u)	3814	53	48-60	81	73-86	26	18-33
12 (r)	921	50	43-57	71	63-78	20	15-27
13 (r)	572	60	52-68	80	72-87	18	13-26
14 (u)	326	43	36-49	73	66-79	29	23-36
Total	8672	53	45-60	78	71-86	24	18-33
p (KW)		<0.001		<0.001		<0.001	

teristics for the Staffelstein Index before and after inpatient treatment as well as for the index increases.

The median direct costs from the heath care insurers' perspective were estimated €2023 for the pooled THA cohort; the direct costs showed a locally significant difference between the departments (Kruskal/Wallis Test P<0.001). Direct costs were highest with a median of €2238 (interquartile range €2126-2238 for this department) and lowest with a median of €1769 (interquartile range €1676-1862 for this department). The median days in rehabilitation ranged along the departments between 19-21 days (median 21 days) and appeared to be rather homogeneous on a global scale (Table III).

Cost-normalized effect estimates (CNEE) for the total cohort were in median 12% Staffelstein Index increase per €1000 investment (department-wise 9-15% per €1000). The 9 departments showed statistically significant differences in the CNEE endpoint (Kruskal/Wallis Test P<0.001, Table III). The lowest CNEE was found for departments 12 and 13 with a median of 9% increase per €1000 investment (respective interquartile ranges 7-12% per €1000 and 6-13% per €1000). Departments 2 and 14 performed with a median of 15% Staffelstein Index increase per €1000 investment (respective interquartile ranges 11-20% per €1000 and 12-19% per €1000).

Socio-demographic cofactors under consideration such as gender or age did not show locally significant association with the department contrasts in the clinical outcome and the CNEE endpoint. For example, in the overall THA cohort a median Staffelstein index increase of 24% was found for both genders (department-wise medians for females 18-33% and for males 18-32%), corresponding to respective median CNEEs of 12% Staffelstein increase per €1000 investment (department-wise medians for both genders 9-16%/€1000).

TKA

In the TKA cohort a median Staffelstein increase of 24% (medians ranging from 17% to 34% between the departments) from a median index of 50% (medians ranging 39-61%) when entering indoor treatment to a median index of 76% (medians ranging 67-92%) at withdrawal was reported. The intra-individual changes differed locally significantly at the 5% level between the departments (Kruskal/Wallis Test P<0.001). In each department, the underlying patients showed notable variation in the primary clinical endpoint: the greatest index change in median was 34% (interquartile range for the respective department's patients 26-43%), the smallest change in median was 19% (patients' interquartile range 13-26%). Table IV summarizes the distribution characteristics for the Staffelstein Index before and after inpatient treatment as well as for the index increases.

The median direct costs from the heath care insurers' perspective were estimated €2023 for the pooled TKA cohort (department-wise medians €1679-2283); the direct costs showed a locally significant difference between the departments (Kruskal/Wallis Test P<0.001). The median days in rehabilitation ranged along the departments between 19-21 days (median 21 days) and again appeared to be rather homogeneous (Table V). Cost-normalized effect estimates (CNEE) for the total cohort were in median 12%

TABLE III.—*THA – health economic outcome.*

Department code (u=urban area r=rural area)	N	Direct costs (days flat [€] x days in rehabilitation [d])	Days in rehabilitation [d]	CNEE [% changes/1.000 €]	
		Median	Median	Median	Q1-Q3
1 (r)	704	1769	19	12	9-16
2 (u)	459	1809	20	15	11-20
3 (r)	521	1836	20	14	11-20
7 (r)	361	2166	21	14	11-18
8 (r)	994	1973	21	13	9-16
11 (u)	3814	2023	21	12	9-16
12 (r)	921	2238	20	9	7-12
13 (r)	572	1977	20	9	6-13
14 (u)	326	1933	20	15	12-19
Total	8672	2023	21	12	9-16
p (KW)		<0.001		<0.001	

TABLE IV.—*TKA - clinical outcome.*

Department code (u = urban area, r = rural area)	N	Pre-SSI: knee [%]		Post-SSI: knee [%]		Change in SSI: knee [%]	
		Median	Q1-Q3	Median	Q1-Q3	Median	Q1-Q3
1 (r)	713	55	46-63	76	68-86	22	17-28
2 (u)	438	39	33-44	68	60-74	28	22-35
3 (r)	564	61	53-71	92	83-94	25	17-34
7 (r)	354	48	40-53	82	74-90	34	26-43
8 (r)	796	48	39-55	75	66-79	26	19-33
11 (u)	3466	51	48-56	78	69-86	26	17-35
12 (r)	1005	46	42-53	67	60-76	20	14-27
13 (r)	504	56	50-65	78	69-83	19	13-26
14 (u)	340	41	34-48	71	63-78	29	22-36
Total	8180	50	44-57	76	67-85	24	17-33
p (KW)		<0.001		<0.001		<0.001	

TABLE V.—*TKA – health economic outcome.*

Department code (u=urban area, r=rural area)	N	Direct costs (days flat [€] x days in rehabilitation [d])	Days in rehabilitation [d]	CNEE: knee [% change/1.000 €]	
		Median	Median	Median	Q1-Q3
1 (r)	713	1769	19	12	9-16
2 (u)	438	1809	20	16	12-20
3 (r)	564	1836	20	14	9-19
7 (r)	354	2166	21	16	12-20
8 (r)	796	1973	21	13	10-17
11 (u)	3466	2023	21	12	8-17
12 (r)	1005	2238	20	9	6-12
13 (r)	504	1977	20	10	7-13
14 (u)	340	1933	20	15	11-19
Total	8180	2023	21	12	8-17
p (KW)		<0.001		<0.001	

Staffelstein Index increase per €1000 investment (department-wise 9-16% per €1000). The 9 departments showed statistically significant differences in the CNEE endpoint (Kruskal/Wallis Test P<0.001).

As already found for the THA cohort, also in the TKA cohort socio-demographic cofactors under consideration again were not found to show locally significant association with the department contrasts in the clinical outcome and the CNEE endpoint.

Discussion

Risk adjustment

The present study contrasted the clinical and economic outcome of 9 competing orthopedic rehabilitation departments in Rhineland-Palatinate by means of a functional assessment of patients after THA and TKA. For this purpose the clinical data of 8672 THA and 8180 TKA patients were collected and a cost/benefit relation estimated based on the achieved clinical improvement and respective direct costs invested by statutory health insurance. For both the THA and TKA cohort a statistically significant as well as a clinically and economically relevant heterogeneity between the rehabilitation departments was found. To clarify the putative influence of socio-demographic variables additional analyzes were performed, which did not resolve the observed heterogeneities in clinical and economic outcomes between the departments under consideration; as a consequence, adjustment for these cofactors was not detailed in the above result presentation.

Relevant outcome determinants in PHCT are not

only age and gender,[13] but also mental state, (chronic) comorbidity, social status, risk behavior of patients may but also the pre-therapeutic function tests before starting treatment have an influence. Of further interest are occurring intercurrent postsurgical complications. However, in this investigation based on completely anonymous patient data, the individual profiles of relevant somatic comorbidity profiles were only made available in terms of aggregate score information: note, that the EVAReha® software only allows for limited documentation of such cofactors; furthermore data protection considerations required the aggregation of the latter to retain the intention of a fully anonymous data set. As then adjustment based on these aggregate comorbidy profile score lacked from sensitivity, the negative findings on association between department contrasts in functional outcome and possibly different background comorbidity patterns between departments need to be reproduced in subsequent investigations based on more detailed individual data.

On the other hand, the results of our study show matching results with the study of Aliyev [14] concerning cohort characteristics, duration of inpatient rehabilitation, the clinical endpoint of change as well as observations on risk adjustment for gender and age. However, the latter investigation was limited to clinical outcome to consider the outcome quality of PHCT; costs or cost-effectiveness were not considered. In addition, the number of cases (N.=141 patients) in the study of Aliyev was rather moderate as compared to the recent study. Comparisons with the American literature are limited due to the shorter inpatient rehabilitation (about 10 to 11 days) and a different organized health care system.[15-17]

Heterogeneities between rehabilitation departments

Statistically significant differences (Kruskal/Wallis $P<0.001$) had to be ascertained between the competing orthopedic rehabilitation departments considering the functional status of patients entering indoor treatment: in particular, the median pre-SSI for THA varied from 42% to 60% between the departments respective for TKA from 39% to 61%. These differences may be explained by possibly different postoperative therapeutic pathways in the acute care hospitals – maybe additionally depending on regional discrepancies: although the introduction of the German DRG (Diagnosis Related Groups) cost calculation standards should have resulted in comparable lengths of stay among the nine rehabilitation departments' acute care clinical partners. Nevertheless the acute care partners cannot be expected to perform identically intensive prerehabilitation preparation for their respective patients. As a consequence the observed heterogeneity in pre-SSI might be an epi-phenomenon of differences in postoperative conservative care standards before admission the rehabilitation units.

Unfortunately no details were available for time period between discharge from acute care hospitals and the beginning of rehabilitation therapy. A delayed admission for rehabilitation is conceivable due to capacity reasons. However in Germany an admission to rehabilitation department should usually take place within three weeks after discharge from acute care. Both the duration of hospital stay and the date of discharge will be recorded in future by developed EVAReha® software.

With regard to heterogeneous results in clinical outcome it has to be mentioned that the nine departments may have varying treatment concepts — detailed evidence-based treatment protocols are currently not available.[12] Furthermore only seven of the nine competing orthopedic rehabilitation departments agreed to provide data on their respective treatment pathway implementations and the underlying intensities of multimodal therapeutic care components. A dedicated evaluation could not be carried out yet due to incompleteness of the data.

Staffelstein Score

The Staffelstein Score [11] was used as disease-specific assessment tool in the present study. Functional assessment in both hip and knee disorders based on the same instrument as requested here to derive parallel benchmarks for each department, is possible by means of several methods: the most common approaches are based on the patient-related WOMAC questionnaire,[18] and the Lequesne Index,[19, 20] or on parallel assessment of the merely function-related Harris Hip Score for THA and on the Knee Society Score for TKA. If, however, the assessment is required to involve both functional and patient-related aspects, the Staffelstein Score can be considered for assessment.[11]

Despite its yet rather limited use, the Staffelstein

assessment's applicability to rehabilitation and its sensitivity to change during short-term functional and patient-related rehabilitation outcome made it a promising candidate for the EVAReha® assessment intentions: both the WOMAC and the Lequesne Index were not originally designed for rehabilitation, but rather for long-term health assessment in the surveillance and therapy of arthritis. The major disadvantage of both assessment instruments is their moderate sensitivity to short-term postoperative changes. In contrast, the Staffelstein Score was specially developed for evaluating short-term postoperative outcome of inpatient rehabilitation.[11, 21, 22] Compared to the WOMAC and Lequesne assessments, both presenting self-assessment questionnaires, the Staffelstein Score also includes an objective functional assessment by the attending physician, which contributes to about 34% of the total score. This combination of both subjective and objective findings may be regarded as an important advantage of Staffelstein assessment. Furthermore, the Staffelstein Score bears high information content for the evaluation of PHCT [14] as well as for the "exploitation of rehabilitation potential".[23] It is now well established in Germany [24] and could be demonstrated in several recent studies as a practically feasible and informative assessment tool.[14, 23, 25] Nevertheless, its major disadvantage of lacking international acceptance and comparability as well as its methodological limitations concerning psychometric determinants must be kept in mind, whenever Staffelstein based longitudinal data are evaluated: its reliability must yet be considered limited, as there is yet no Standard Operation Procedure available for diagnostic assessments by means of the Staffelstein Score. Furthermore its validity has not been proven convincingly, but rather been postulated due to its derivation from previously psychometrically evaluated instruments. On the other hand, this weakness concerning both validity of Staffelstein based assessments and maybe even objectivity of the latter — note that functional information is entered by the physician in charge and thereby underlies the unavoidable bias mechanisms of self-reported physician data in unblinded assessments — must be expected systematic and similar for all 9 rehabilitation departments considered in this investigation; as a consequence, for the sake of external benchmarking based on inter-department contrasts in Staffelstein based endpoints, the contrasts between departments shall have remained unbiased. Nevertheless, a conflict of interest for the physicians and data-assessors cannot be denied due to the fact they are employees of the rehabilitation departments participating in this benchmarking study.

In summary, the Staffelstein data provided by the EVAReha® assessment combines clinical findings from the practitioner's perspective as well as patient information in three dimensions targeting on pain, ADL (activities of daily living) and joint function. Nevertheless, no explicit data on patient´s quality of life were collected before and after PHCT by means of a questionnaire such as the EuroQol-5D or the SF-12 instruments. For this purpose, the EVAReha® software will be continuously developed; in the near future the implementation of quality of life assessments can be expected.

In general, the benchmark is not restricted to the Staffelstein Index. Although the Staffelstein instrument is rather limited because of its German-only available version, this questionnaire is quite similar to the internationally accepted Harris Hip Score (THA) and imitates this score for TKA patients. As a consequence the derived cost effectiveness profiles can be compared to Harris Hip Score related cost benefit estimates and are therefore not thoroughly restricted to our scenario. Furthermore it should be mentioned, that the EVAReha® software is only a technical implementation of the Staffelstein assessment and therefore does not imply any specific trial or outcome impact due to its use in data documentation.

Model assumptions for health economic evaluation

In order to imitate a utility scale (ranging from -100 to +100%) the Staffelstein Score was linearly transformed into an index range. This does not affect internal validity of the clinical results, but allows for a simpler cost effectiveness parameterization when being related costing data.[26, 27] The health economic target parameter used in this investigation, CNEE, corresponds to cost-effectiveness analysis and is derived as the inverse of the more well-known Cost Effectiveness Ratio (CER), *i.e.* the individual cost / benefit ratio. However, since the intra-individual Staffelstein Index changes also allow for negative values (in the sense of an aggravation of patient-condition during inpatient rehabilitation), assessment of CER medians would be severely biased. To

retain correct interpretation of negative individual cost effectiveness values, the CER inverse, denoted as cost-normalized effect estimate (CNEE), was applied in this investigation. Despite its methodological advantages, however, the CNEE is yet rarely used and therefore does not allow for direct comparisons of CNEE based contrasts with CER contrasts in the literature.

Furthermore, the CNEE suffers from the same weaknesses as any other cost effectiveness parameterization, as it depends on the validity of both clinical outcome and (direct) cost estimates. In the recent setting, the latter are determined by the information on day rates for each department and the individual duration of stay in the respective rehabilitation unit. The median duration of stay for all departments was 21 days, so this factor can be considered of moderate variation and model impact. The daily rate of the departments, however, differed in a somewhat more relevant order: they result from individual negotiations of the rehabilitation department with the statutory health insurance. As a consequence, these daily cost rates include not only the costs of human resources, but also possible interests of the respective clinics, for example, concerning investment plans for new buildings, renovations, etc. However, increases in payment rates are limited by the government on the annual increase in the gross basic wage sum. In summary, the cost rates do not necessarily reflect the respective rehabilitation departments' service offers or therapy concept volumes, but must rather be considered as a surrogate of business negotiation. This might explain the findings contained in Tables III, V concerning the heterogeneous cost effectiveness profiles among departments: department 12, the one with the highest daily cost rate of €111.92 per day, showed one of the worst median outcome profiles both in the THA and the TKA cohort, corresponding to one of the two worst CNEE estimates (median CNEE THA and TKA 9% Staffelstein Index increase per €1000). On the other hand, department 2 presented the lowest day rate of €90.45 per day and a clinical outcome far better than found in median for the pooled THA and pooled TKA cohorts, respectively, corresponding to the best observed results in terms of the CNEE parameterization (THA 15% and TKA 16% Staffelstein Index increase per €1000). However, department 7 presenting best median clinical outcome and the second highest daily cost rate of €103.13 per day, only reached a place in the top third of the CNEE based rankings (Tables III, V).

Health economic benchmarking?

In summary, these examples demonstrate, that the CNEE based benchmarking process introduces a new perspective in benchmarking: the "best ranked" competitor is yet no longer the one with the best clinical outcome profiles, but is rather the competitor with "optimum return to investment" regarding the benefit cost ratio of clinical outcome and financial investment. Therefore health service providers performing at a higher level of cost efficiency with consistently satisfying outcome profiles may turn out with "better" ranks than overly expensive clinical top performers. However, application of this extended benchmark concept must thoroughly bear in mind the fact, that a clinical outcome parametrization such as chosen here also depends on the underlying patient cohort's baseline profiles: if the (median) baseline Staffelstein index is observed smaller for a patient, this may imply greater potential for improvement (and therefore greater potential for a "large" clinical outcome median in the CNEE computation), but may also present evidence for a more severe comorbidity (yet hardly representable in daily cost rates of the respective department!). As a consequence, adjustment for relevant influence factors becomes overly crucial for this extended benchmarking approach based on a health economic parametrization.

In addition, it should be mentioned that the numbers of cases in rehabilitation facilities varied notably in the recent investigation. For example, department 14, which provided the best results for the clinical and economic endpoints for both THA and TKA, documented the lowest number of cases, whereas department 11 having documented a total of 3.814 THA and 3.466 TKA patients (Table I).

Conclusions

External comparative quality management in German hospitals is mandatory since 2004. Every two years the results are published in form of a structured quality report,[28] and it has gained increasing importance in health care politics over the past decade.

On the other hand, benchmarking data such as the EVAReha® data presented in this investigation can also be applied by the participating departments to derive an internal benchmark: screening the acquisition years 2007, 2008 and 2009 in terms of the CNEE parameterization, for example department 2 could show monotonic improvement over time by THA cohort medians of 13%, 17% and 18% per €1000 Staffelstein Index increase in the respective years as well as corresponding TKA cohort medians of 12%, 17% and 18% per €1000. During the same period, department 14, having started 2007 with the best results, showed a monotonic decrease in CNEE medians over the three years with THA cohort medians of 17%, 14% and 12% per €1000, respectively, and TKA cohort medians of 18%, 15% and 12% per €1000, respectively. If these information are provided regularly to the respective department, the demonstration of recent achievement potential can be expected overly appreciated.[29]

The results of annual evaluations of the EVAReh® a database is communicated to the participating departments biannually according to a standardized method protocol (statistical analysis and reporting) by the Medical Service of the Health Insurer Rhineland-Palatinate. The external benchmark valuations, however, remain blinded between departments. In addition a benchmark information on cost effectiveness as well as on time trends in the outcome parameters were not yet provided, but would be desirable: on one hand, it would enable departments to learn from the "best" and to implement corresponding peer cooperations, to then identify own shortcomings and potentials, and thereby to analyze and optimize local processes.[9] On the other hand, the health insurances might use the results in subsequent negotiation processes on daily cost rates and for department selection or at least recommendation algorithms to be communicated to their insured. The economic and political impact of possible consequences of benchmarking such as presented here shall motivate the yet limited acceptance of the overall approach.

References

1. Gehrke W, Arnold W. [Mobility outcome after knee prosthesis implantation and subsequent physical therapy--factors influencing rate of success]. Die Rehabilitation 2001;40:156-64 [Article in German].
2. Deutsche Arthrose Hilfe e.V.: DAH; [Internet]. Available from: http://www.arthrose.de [cited on 2015, Nov 11].
3. Schnabel P, Borelli S. Endoprothesenregister - Höchsten Anforderungen genügen. Dtsch Arztebl 2011;108:A-2598-A2602.
4. Kurtz S, Ong K, Lau E, Mowat F, Halpern M. Projections of primary and revision hip and knee arthroplasty in the United States from 2005 to 2030. J Bone Joint Surg Am 2007;89:780-785.
5. EPRD: EPRD - Endoprothesesenregister Deutschland [Internet]. Berlin: EPRD Deutsche Endoprothesenregister gGmbH [cited 2015 nov 11]. Available from: http://www.eprd.de/.
6. Deutscher Verlag für Gesundheitsinformation GmbH: Operationen-Endoprothetik.de [Internet]. Heidelberg: Deutscher Verlag für Gesundheitsinformation GmbH [cited 2015 Nov 11]. Available from: http://www.operation-endoprothetik.de/.
7. Rupp S, Wydra G. [Rehabilitation following total knee replacement. Conservative orthopedics and sport science]. Orthopade 2012;41:126-35.
8. DESTATIS: Grunddaten der Vorsorge- oder Rehabilitationseinrichtungen 2012 [Internet]. Wiesbaden: Statistisches Bundesamt [cited 2015 Nov 11]. Available from: https://www.destatis.de/DE/Publikationen/Qualitaetberichte/Gesundheitswesen/Grunddatenrehaeinrichtungen.pdf.
9. Bredl K, Husig S, Angele MK, Luring C. [Benchmarking in the clinical arena. A potential answer to the dynamic changes in the health care system]. Orthopade 2010;39:785-91 [Article in German].
10. MDK Rheinland-Pfalz. EVA-Reha; [Internet]. Alzey: Medizinischer Dienst der Krankenversicherung Rheinland-Pfalz [cited 2015; Nov 11]. Available from: http://www.mdk-rlp.de//eva-reha/.
11. Middeldorf S, Casser H. Verlaufs- und Ergebnisevaluation stationärer Rehabilitationsmaßnahmen nach alloarthroplastischem Hüft- und Kniegelenkersatz mit dem Staffelstein-Score. Orthopädische Praxis 2000;36:230-8.
12. Di Monaco M. Rehabilitation after hip and knee arthroplasty: where are we now? Work in progress to build up evidence-based protocols. Eur J Phys Rehabil Med 2013;49:875-6.
13. Vincent HK, Alfano AP, Lee L, Vincent KR. Sex and age effects on outcomes of total hip arthroplasty after inpatient rehabilitation. Arch Phys Med Rehabil 2006;87:461-7.
14. Aliyev RM. [Total hip replacement using the Staffelstein score: outcome of inpatient rehabilitation]. Orthopade 2010;39:1163-70 [Article in German].
15. Munin MC, Rudy TE, Glynn NW, Crossett LS, Rubash HE. Early inpatient rehabilitation after elective hip and knee arthroplasty. JAMA 1998;279:847-52.
16. Vincent KR, Vincent HK. A multicenter examination of the Center for Medicare Services eligibility criteria in total-joint arthroplasty. Am J Phys Med Rehabil 2008;87:573-84.
17. Vincent KR, Vincent HK, Lee LW, Weng J, Alfano AP. Outcomes after inpatient rehabilitation of primary and revision total hip arthroplasty. Arch Phys Med Rehabil 2006;87:1026-32.
18. Bellamy N, Buchanan WW, Goldsmith CH, Campbell J, Stitt LW. Validation study of WOMAC: a health status instrument for measuring clinically important patient relevant outcomes to antirheumatic drug therapy in patients with osteoarthritis of the hip or knee. J Rheumatol 1988;15:1833-40.
19. Lequesne MG, Mery C, Samson M, Gerard P. Indexes of severity for osteoarthritis of the hip and knee. Validation--value in comparison with other assessment tests. Scand J Rheumatol Suppl 1987;65:85-9.
20. Ludwig FJ, Melzer C, Grimmig H, Daalmann HH. [Cross cultural adaptation of the lequesne algofunctional indices for german speaking patients with osteoarthritis of the hip and the knee]. Die Rehabilitation 2002;41:249-57 [Article in German].

21. Slupik A, Bialoszewski D. Comparative analysis of clinical usefulness of the Staffelstein Score and the Hospital for Special Surgery Knee Score (HSS) for evaluation of early results of total knee arthroplasties. Preliminary report. Ortop Traumatol Rehabil 2007;9:627-35.
22. Slupik A, Bialoszewski D. A comparative analysis of the clinical utility of the Staffelstein-score and the hospital for special surgery knee score (HSS) in monitoring physiotherapy of total knee replacement patients--preliminary study. Ortop Traumatol Rehabil 2009;11:37-45.
23. Heisel J. [Rehabilitation following total hip and knee replacement]. Der Orthopade 2008;37:1217-32.
24. Claes L KP, Perka C, Rudert M. AE-Manual der Endoprothetik: Hüfte und Hüftrevision. Heidelberg: Springer-Verlag; 2011.
25. Torbati T, Schladitz G. Verlaufs- und Ergebnisevaluation stationärer Rehabilitationsmaßnahmen nach alloarthroplastischem Hüftgelenkersatz mit dem Staffelstein-Score. Orthopädische Praxis 2001;37:236-42.
26. Krauth C, Hessel F, Klingelhöfer HE, Schweikert B, Hansmeier T, Wasem J. [Health economic evaluation of rehabilitation programmes in the „rehabilitation sciences" research funding programme in Germany]. Die Rehabilitation 2005;44:297-306.
27. Schulenburg MGW. Gesundheitsökonomik. Tübingen: Verlag Mohr Siebeck; 1999.
28. Friedemann J, Schubert HJ, Schwappach D. [On the comprehensibility of German hospital quality reports: systematic evaluation and need for action]. Gesundheitswesen 2009;71:3-9.
29. Mertins K KH ed. Benchmarking, Leitfaden für den Vergleich mit den Besten. Düsseldorf: Symposium Publishing; 2008.

The data presented in this manuscript are contained in the Doctoral Thesis of Mrs M Grams submitted to Witten / Herdecke University, Faculty of Health, in March 2013.

Some results were presented at the 59th annual Meeting of the German Society for Medical Computer Science, Biometry and Epidemiology (GMDS), Göttingen, September 2014.

Acknowledgements.—The authors gratefully thank the "German Arthrosis Help e.V" for providing the research grant for Mrs Mashiba Grams. The authors are grateful to Dr. Manfred Nosper for giving us the opportunity to perform the recent evaluations by means of the EVAReha® software and the data acquired by means of this software. Finally, the authors thank the rehabilitation departments participating in the present study for providing the clinical outcome data.

Funding.—The "German Arthrosis Help e.V." provided a research grant for Mrs M Grams to enable the data processing and analysis (research grant: P208-A144-Krummenauer-EP1-allgemeinen-II-12k-2009-10).

Conflicts of interest.—The authors certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

Received on November 4, 2014.
Accepted for publication on May 21, 2015.
Epub ahead of print on May 26, 2015.