

Comparative effectiveness research across two spine registries

Emin Aghayev · Julia Henning · Everard Munting ·
Peter Diel · Patrick Moulin · Christoph Röder ·
On behalf of the SWISSspine and Spine Tango Registry groups

Received: 21 November 2011 / Revised: 25 January 2012 / Accepted: 1 March 2012 / Published online: 14 March 2012
© Springer-Verlag 2012

Abstract

Background Comparative effectiveness research in spine surgery is still a rarity. In this study, pain alleviation and quality of life (QoL) improvement after lumbar total disc arthroplasty (TDA) and anterior lumbar interbody fusion (ALIF) were anonymously compared by surgeon and implant. **Methods** A total of 534 monosegmental TDAs from the SWISSspine registry were analyzed. Mean age was 42 years (19–65 years), 59 % were females. Fifty cases with ALIF were documented in the international Spine Tango registry and used as concurrent comparator group for the pain analysis. Mean age was 46 years (21–69 years), 78 % were females. The average follow-up time in both samples was 1 year. Comparison of back/leg pain alleviation and QoL improvement was performed. Unadjusted and adjusted probabilities for achievement of minimum clinically relevant improvements of 18 VAS points or 0.25 EQ-5D points were calculated for each surgeon.

Results Mean preoperative back pain decreased from 69 to 30 points at 1 year (Δ 39pts) after TDA, and from 66 to

27 points after ALIF (Δ 39pts). Mean preoperative QoL improved from 0.34 to 0.74 points at 1 year (Δ 0.40pts). There were surgeons with better patient selection, indicated by lower adjusted probabilities reflecting worsening of outcomes if they had treated an average patient sample. ALIF had similar pain alleviation than TDA.

Conclusions Pain alleviation after TDA and ALIF was similar. Differences in surgeon's patient selection based on pain and QoL were revealed. Some surgeons seem to miss the full therapeutic potential of TDA by selecting patients with lower symptom severity.

Keywords Comparative effectiveness · Spine registry · SWISSspine · Total disc arthroplasty · Benchmark

Abbreviations

ALIF	Anterior lumbar interbody fusion
COMI	Core outcome measures index
EQ-5D	EuroQoL-5D instrument
MCRIQL	Minimum clinically relevant improvement of quality of life = 0.25 EQ-5D points [2]
MCRPI	Minimum clinically relevant pain improvement = 18 VAS points [10]
NASS	North American Spine Society outcome assessment instrument
QoL	Quality of life
TDA	Total disc arthroplasty
VAS	Visual analogue scale

E. Aghayev (✉) · C. Röder
Institute for Evaluative Research in Medicine,
University of Bern, Stauffacherstrasse 78,
3014 Bern, Switzerland
e-mail: emin.aghayev@memcenter.unibe.ch

J. Henning · P. Diel
Inselspital, University Hospital, University of Bern,
Bern, Switzerland

E. Munting
Clinique Saint Pierre, Avenue Reine Fabiola 9,
1340 Ottignies, Belgium

P. Moulin
Swiss Paraplegic Center Nottwil, Guido A. Zäch Strasse 1,
6207 Nottwil, Switzerland

Introduction

In 2010, the Obama administration provided 1.1 billion USD for so-called comparative effectiveness research.

Besides comparison of different therapies, comparison of implants and even of health care providers is possible. “Treatment success”, enabling comparison of more or less effective therapies, may be influenced by a multitude of factors and is rarely clearly defined. A minimum clinically relevant improvement of a certain symptom could be used for its definition, but also the average improvement of the same symptom within a group of implants or physicians, a value referred to as “benchmark”.

Recently, the possibility of benchmarking between single lumbar disc prosthesis against the pool of all other prostheses was shown in the SWISSspine registry [1–3]. The example of this governmentally mandated registry with its uniform documentation opens benchmarking possibilities between implants and surgeons, but the lack of an included comparator like anterior lumbar interbody fusion (ALIF), hampers conclusions about general superiority or inferiority of lumbar TDA. Therefore, we drew information from the international Spine Tango registry for enabling comparative effectiveness research between lumbar TDAs, ALIF and the healthcare providers.

The current study therefore anonymously compared back and leg pain alleviation after total disc arthroplasty and ALIF stratified by implant and surgeon from the SWISSspine and Spine Tango registries. We hypothesized that TDA was not inferior to ALIF regarding back and leg pain alleviation and that there were significant differences in outcomes of pain alleviation and quality of life improvement between the best and the worst performing TDA surgeons.

Materials and methods

SWISSspine registry

The structure and setup of the registry was described elsewhere [1, 2]. Currently, 49 surgeons from 32 hospitals are contributing data since 2005. In November 2010, there were data on 534 monosegmental TDAs with a preoperative and at least one post-operative NASS and EQ-5D questionnaire available from 313 females (59 %) and 221 males (41 %). Mean age was 42 years (range 19–65 years).

From the eight different disc prosthesis models, *four* had at least ten documented cases. They contributed a total of 521 cases (on average ~130 cases per prosthesis, range 37–247). Thirteen remaining cases had disc prostheses from other suppliers and were grouped for further comparisons (“other TDA prostheses”).

For synchronization of follow-up times, the last available follow-up per patient until the end of the second post-operative year was used for statistical assessment. The average follow-up time in this SWISSspine sample was 1 year.

Fifteen of 49 surgeons had 10 documented cases or more each. These 15 surgeons covered 418 cases (average 27 cases per surgeon, range 10–76). The remaining 116 patients treated by the remaining 34 surgeons (average 3 cases per surgeon, range 1–9) were grouped for further comparisons (“other” surgeons).

The 15 individual surgeons and the group of 34 “other” surgeons on the one hand and 4 implant suppliers and the group of “other TDA prostheses” on the other hand were compared in an anonymised way regarding patient back and leg pain alleviation and quality of life improvement.

Spine Tango registry

The registry of the Spine Society of Europe allows documentation of different spinal surgical and conservative procedures [4–7] and has currently a case load of over 40,000 surgeries. Being a non-mandatory registry, the users are asked to document primary and follow-up surgeon-based forms and at least one pre- and post-operative patient-based COMI form [8].

Comparator group from Spine Tango

The fusion of affected lumbar segments in surgical candidates with chronic low back pain has been the standard surgical procedure for almost 50 years and remains the gold standard until today [9]. The inclusion criteria for the treatment reference group (ALIF) were the following: monosegmental procedure + linkage to at least one completed preoperative and one post-operative COMI form + lumbar/lumbo-sacral level of procedure + degenerative disease as main diagnosis + no previous surgery on the same level + retroperitoneal/transperitoneal approach + anterior fusion between adjacent vertebral bodies + rigid stabilization using a cage. The query resulted in 50 single patients from 3 surgeons (average 16 cases per surgeon, range 12–22). The female/male ratio was 39/11. Mean age was 46 years (range 21–69 years). The last available follow-up per patient within 2.5 years was used for the statistical assessment. The average follow-up time in the sample was 1 year. The 50 selected patients had no documented EQ-5D forms and the comparison between TDA (SWISSspine) and ALIF (Spine Tango) was therefore limited to post-operative pain alleviation.

Statistical analysis

We compared probabilities for achievement of minimum clinically relevant pain (MCRPI) and quality of life (MCRIQL) improvement of 18 VAS points [10] and of 0.25 EQ-5D points [2]. Preoperative pain levels influence post-operative pain alleviation and similarly, preoperative

quality of life influences its post-operative improvement [1, 2]. Further analyzed co-variables, such as implant, surgeon, depression, age, gender, follow-up interval and length of hospitalization had no significant influence on pain alleviation. Therefore, in a first step a univariate logistic regression (MRCPI or MCRIQL vs. implant or vs. surgeon) resulting in non-adjusted probabilities and, in a second step, a generalized linear model (MRCPI or MCRIQL vs. implant or vs. surgeon) adjusted by preoperative pain level or quality of life were calculated.

For within-group comparisons, Wilcoxon signed-rank or Chi-square test was used. For between-group comparisons, Wilcoxon rank-sum test was used. α was set to 0.05. All statistical analyses were conducted using SAS 9.2 (SAS Institute Inc, USA).

Results

Comparison of both samples only showed significant differences regarding patient age and gender. ALIF patients were on average 4 years older and there were 19 % more females. Pre-operative and post-operative pain values as well as pre- to post-operative changes were not significantly different between the samples.

TDA outcomes

Mean preoperative back pain on VAS was 69 points and leg pain was 54 points and mean post-operative back pain was 30 points and leg pain was 23 points (pre- to post-op,

both $p < 0.001$). Hence, there was a back pain alleviation of 39 points and leg pain alleviation of 31 points at the last available follow-up. Mean preoperative quality of life was 0.34 points on EQ-5D and improved to 0.74 points at the last follow-up (pre- to post-op $p < 0.001$). The improvement was equal to 0.4 points.

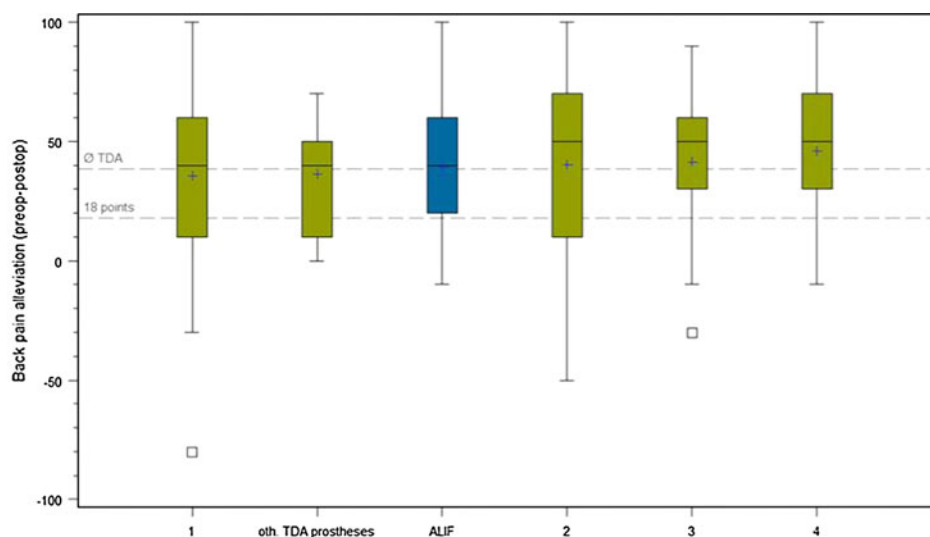
ALIF outcomes

Mean preoperative back pain on VAS was 66 points and leg pain was 49 points and mean post-operative back and leg pain were 27 and 22 points, respectively (pre- to post-op, both $p < 0.001$). Back pain alleviation of 39 points and leg pain alleviation of 27 points were observed.

Stratification by supplier

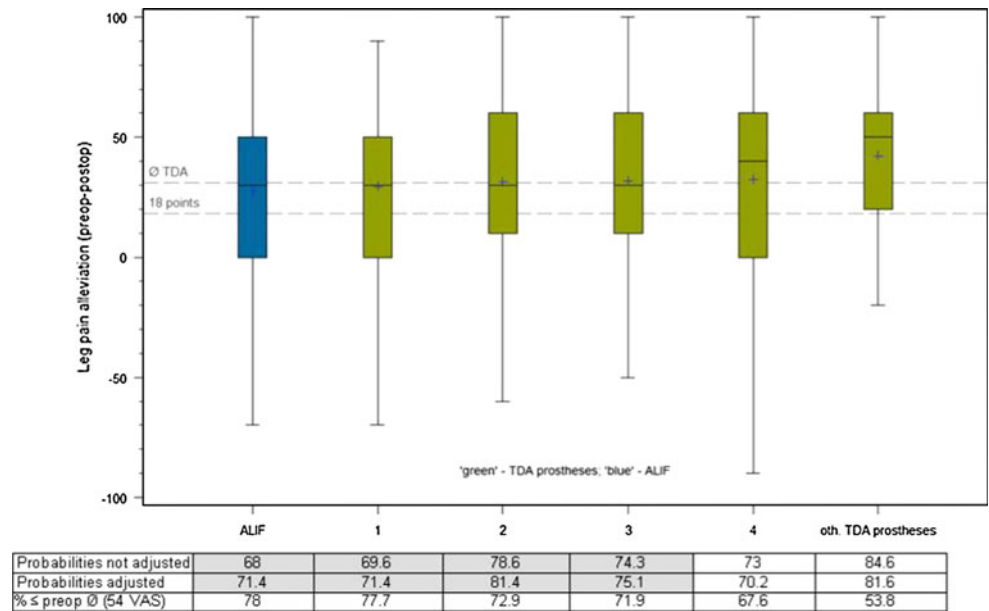
Figures 1 and 2 show similar average and median values for the post-operative back and leg pain alleviation stratified by disc prosthesis model in the SWISSspine and Spine Tango groups. The difference between the best and the worst average back pain alleviation was only 9 VAS points (Fig. 1). The variation of leg pain alleviation between major disc models was relatively low, though the grouped “other TDA prostheses” were better than ALIF by an average of 15 VAS (n.s.) (Fig. 2). For all four major prosthesis suppliers in SWISSspine as well as for ALIF in Spine Tango, the post-operative pain alleviation was approximately twice the MCRPI for back pain and slightly less than twice the MCRPI for leg pain.

Fig. 1 Back pain alleviation after TDA (green) and after ALIF (reference, blue) by supplier. The upper reference line represents the average post-operative back pain alleviation of 39 points. The lower reference line represents the 18 VAS points regarded as minimum clinically relevant pain improvement. The increased adjusted probabilities are marked in grey



Probabilities not adjusted	73.7	69.2	78	74.3	81.4	78.4
Probabilities adjusted	78.7	73	83	74.2	85	76.2
% with preop ≥ 43.8 VAS	39.7	46.2	42	50.3	51.4	54.1
% \leq preop 0 (69 VAS)	82.6	84.6	86	72.5	78.6	70.3

Fig. 2 Leg pain alleviation after TDA (green) and after ALIF (reference, blue) by supplier. The upper reference line represents the average post-operative back pain alleviation of 31 points. The lower reference line represents the 18 VAS points regarded as minimum clinically relevant pain improvement. The increased adjusted probabilities are marked in grey



Unadjusted and adjusted probabilities

If a prosthesis model gains probabilities of providing the MCRIP after adjustment for preoperative pain or quality of life, then the surgeons have implanted it into patients with lower preoperative pain or higher quality of life levels than the average, the benchmark. This is the case for three prosthesis models and the ALIF group (Fig. 1). If prosthesis loses probabilities for providing minimum clinically relevant pain alleviation after adjustment, the surgeons using it have applied stricter inclusion criteria than the benchmark in terms of preoperative pain or quality of life. This is also reflected by the proportion of cases with preoperative back pain levels above a previously revealed threshold level of 43.8 points on VAS [2] which is higher in the relatively more successful prosthesis models, or by the higher proportion of patients with preoperative back pain levels below the benchmark of 69 VAS points in the relatively less successful models. All prosthesis models with theoretical outcome improvement after adjustment are highlighted in grey.

There is a similar picture for leg pain, except for the ALIF comparator group performing slightly worse regarding leg pain alleviation. Given the non-significant and clinically irrelevant differences, and considering the fact that patients were on average 4 years older, which the model was not adjusted for, the ALIF performance is well comparable with that of TDA (Fig. 2).

Regarding improvement of quality of life, the four supplier’s products were similar in their outcomes and the grouped other prostheses were even slightly better (Fig. 3). The difference between the worst and best average

improvement of quality of life was 0.16 points (n.s.). Once again, three prosthesis models had better adjusted probabilities, i.e. their outcomes could have theoretically been better had all their patients had at least the average preoperative quality of life of 0.342 points EQ-5D or less.

Stratification by surgeon

The stratification of back and leg pain alleviation by surgeon showed more variation in outcomes (Figs. 4, 5). With one exception, the average back pain alleviation of all surgeons was above the MCRPI of 18 VAS points. For leg pain alleviation, there were 2 TDA and 1 ALIF surgeons below the MCRPI. There are surgeons with a superior performance showing more than 75 % of the patients with achieved MCRPI (Fig. 4, lower reference line) and some do even show outcomes where around 75 % of the patients have an over-average post-operative back pain alleviation (Fig. 4, upper reference line).

There were seven surgeons (1 ALIF) who strictly selected patients regarding preoperative back pain and eight surgeons (1 ALIF) who selected patients similarly strict regarding leg pain. This is indicated by the lower adjusted than the non-adjusted probabilities reflecting worsening of patient outcome if an average patient would have been treated. Other surgeons had higher adjusted probabilities indicative of lower preoperative pain values in their patient sample compared with an average patient pain level (Figs. 4, 5).

Further to the right on Fig. 4, more patients with a minimum preoperative back pain level of at least 43.8 VAS points (acc. to the recommended threshold [2]) can be

Fig. 3 Improvement of quality of life after TDA by supplier. The *upper reference line* represents the average improvement of quality of life of 0.396 EQ-5D points. The *lower reference line* represents the 0.25 points regarded as minimum clinically relevant quality of life improvement. The increased adjusted probabilities, which means better outcome if the disc implant would be implanted in an average patient, are marked in *grey*

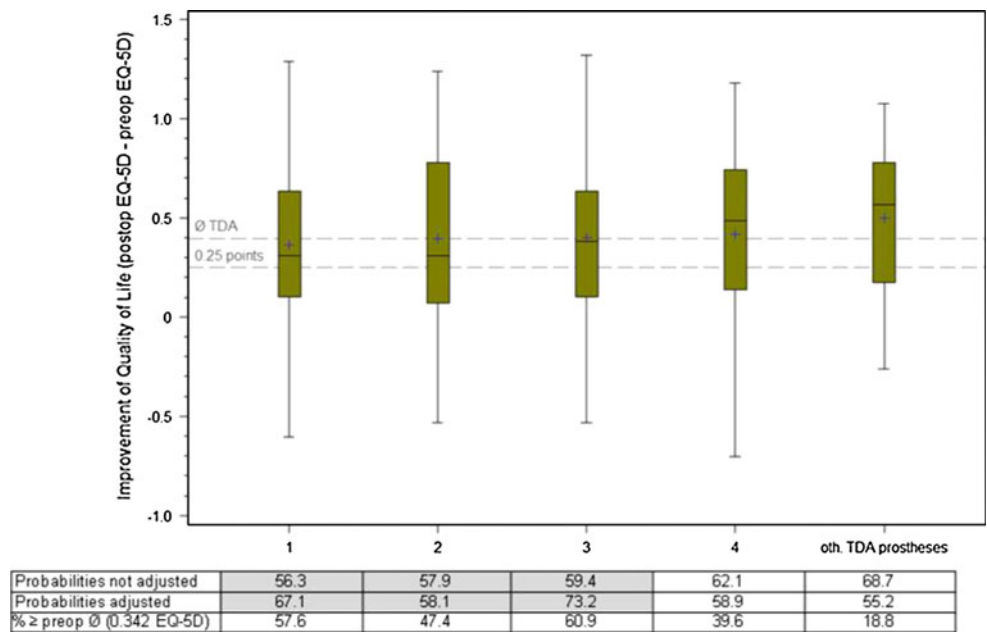
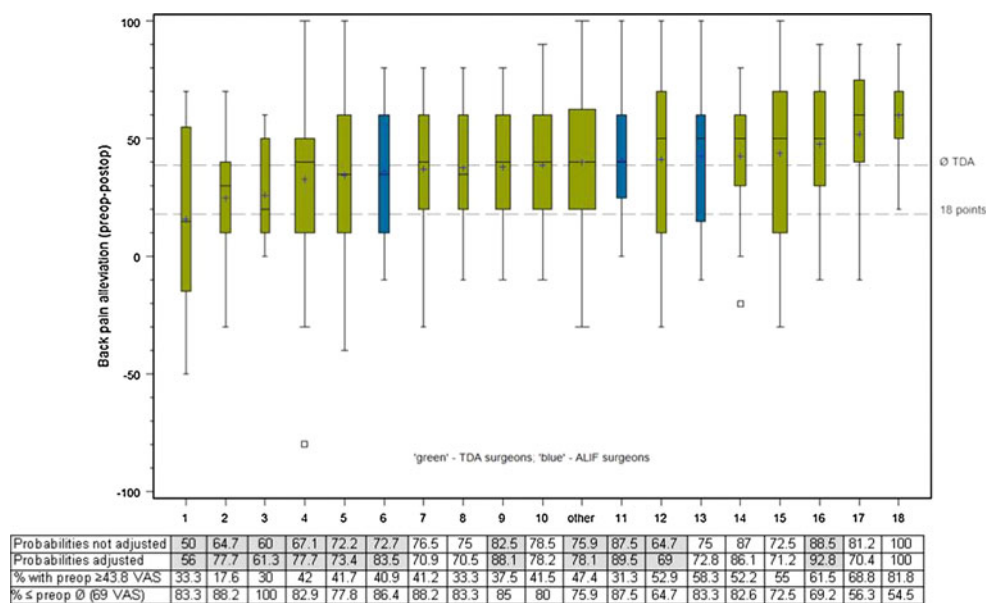


Fig. 4 Back pain alleviation after TDA (*green*) and after ALIF (comparator, *blue*) by different surgeons. The *upper reference line* represents the average post-operative back pain alleviation of 39 points. The *lower reference line* represents the 18 VAS points regarded as minimum clinically relevant pain improvement. The thicknesses of box plots vary proportionally to the respective subgroup size. The increased adjusted probabilities are marked in *grey*



observed (table in Fig. 4). Similarly, towards the right, fewer patients under the average preoperative back pain level of 69 VAS points were observed (table in Fig. 4).

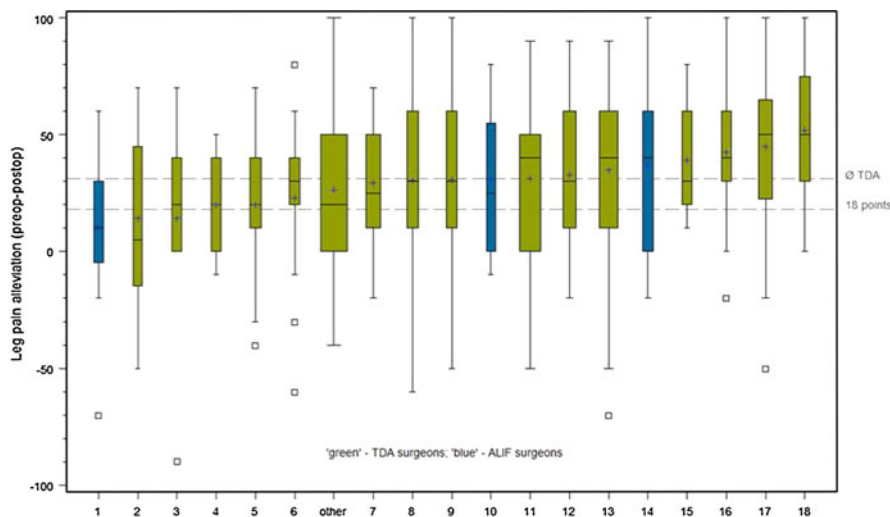
Post-operative improvement of quality of life also had more variation than stratification by supplier (Fig. 6). Ten out of 16 surgeons had higher adjusted probabilities reflecting an existing potential for further post-operative increase of quality of life (Fig. 6). Surgeon 2 for example had almost 77 % of the patients with preoperative quality of life above the average in the registry and the probability of achieving a MCRIQL of 0.25 points on EQ-5D was only 29 %. For an average patient in the pool, a 48 % probability may be expected. The surgeon with the best patient

selection for this outcome was number 16, who had only 6 % of the patients above the preoperative average quality of life and a 82 % non-adjusted and a 52 % adjusted probability for reaching the MCRIQL.

Discussion

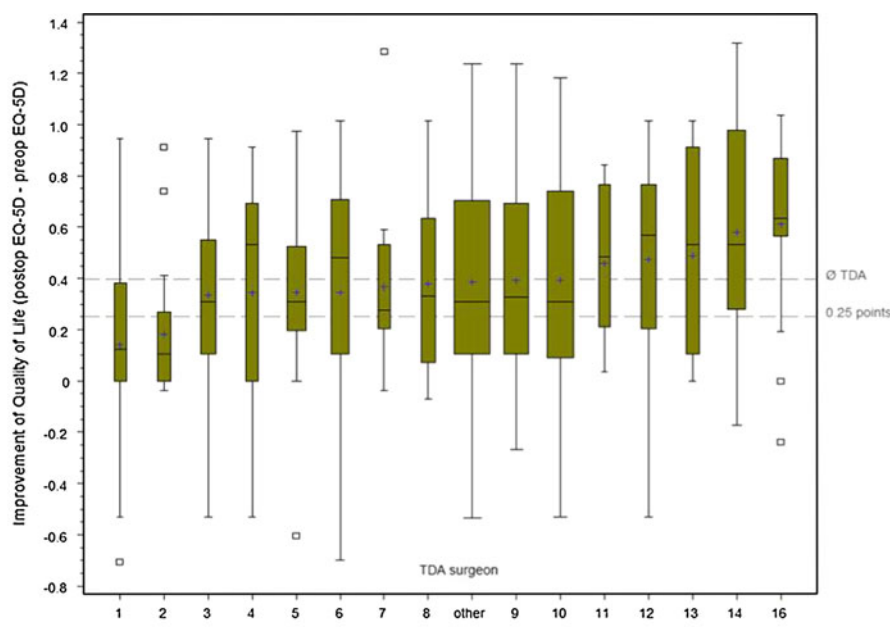
Our first hypothesis was confirmed in that some TDA implant types had a slightly better back pain alleviation, some were slightly worse than with ALIF. Regarding leg pain relief, all TDA prostheses achieved slightly higher values. None of the differences were significant and the

Fig. 5 Leg pain alleviation after TDA (green) and after ALIF (comparator, blue) by different surgeons. The upper reference line represents the average post-operative leg pain alleviation of 31 points. The lower reference line represents the 18 VAS points regarded as minimum clinically relevant pain improvement. The thicknesses of box plots vary proportionally to the respective group size. The increased adjusted probabilities are marked in grey



Probabilities not adjusted	56.3	41.7	58.3	70	60.9	76.5	69	80	80.8	70.6	75	64.5	72.2	75.4	72.7	90.9	88.2	85	93.7
Probabilities adjusted	65.6	39.8	57	75.1	63.4	83	73.2	84.5	85.1	70.7	79	64.4	69.5	72.3	72.6	92.6	88.9	82.5	92.6
% \geq preop \emptyset (54 VAS)	93.8	83.3	91.7	100	82.6	88.2	79.3	77.5	69.2	64.7	75	77.6	72.2	66.2	68.2	63.3	70.6	55	56.3

Fig. 6 Improvement of quality of life after TDA by different surgeons. The upper reference line represents the average improvement of quality of life of 0.396 EQ-5D points. The lower reference line represents the 0.25 points regarded as minimum clinically relevant quality of life improvement. The thicknesses of box plots vary proportionally to the respective group size. The increased adjusted probabilities are marked in grey



Probabilities not adjusted	33.3	29.4	63	61.5	67.5	55.6	54.5	52.9	56.4	55.4	56	72.7	69.6	52.9	76.9	82.4
Probabilities adjusted	31.5	48	82.6	31.8	88	46	72.7	40.2	65.3	58.2	65.8	85.5	62.3	53.4	80.8	51.6
% \geq preop \emptyset (0.342 EQ-5D)	58.3	76.5	74.1	15.4	72.5	38.9	63.6	35.3	53.9	47.7	54.7	63.6	30.4	47.1	38.5	5.9

ALIF sample was 4 years older. The TDA outcomes, pain relief and quality of life improvement within a sample of skilled and certified spine surgeons were not significantly different and we had to consequently reject our second hypothesis, but average differences in pre- to post-operative pain levels and EQ-5D scores were sufficiently large to be regarded as clinically relevant.

Total disc arthroplasty being a relatively new treatment method has gained wide use in developed countries. Comparisons of different disc implants in the literature are largely missing as practically all previous studies on total disc arthroplasty report on only one implant or compare

two implants regarding one specific issue. The study designs of these investigations are frequently quite different, resulting in difficult and inaccurate comparisons. Collection of nationwide data on disc arthroplasty and use of standard documentation instruments allowed for reasonable comparisons between different implants and surgeons.

According to the published earlier results, the major implants did not have any significant influence on post-operative functional outcome [2]. Our comparative effectiveness analysis showed that the four most frequently used prosthesis models provide good post-operative back and

leg pain alleviation and similar improvement of quality of life. Presented results support our assumption that the post-operative outcome differences rather lie in other factors, such as the surgeon and his patient selection.

Our recent study revealed probable differences in patient selection resulting in outcome variations [3]. These differences were confirmed in the present study. There was considerable variation of the results when stratified by surgeon. According to common knowledge treatment outcome is influenced by its respective preoperative status [1, 2]. Some surgeons seem to miss the full pain alleviation and quality of life improvement potential by selection of patients with rather low preoperative pain levels and a rather good preoperative quality of life.

A major drawback of the SWISSspine registry is the absence of a comparator for lumbar TDA which could e.g. be ALIF. Therefore, we used ALIF data from the international spine registry Spine Tango [4–6, 11]. Our results showed that ALIF patients had the same back pain alleviation as TDA patients. Also, leg pain alleviation after ALIF was not significantly different from that after TDA, though ALIF patients had generally slightly lower preoperative pain levels. Stratification by surgeon also showed variation of pain alleviation in ALIF surgeons. Two of the three ALIF surgeons apparently also miss the full potential of pain alleviation and quality of life increase by selection of patients with low preoperative pain levels.

The above observations underline the importance of making a good indication for surgery. Within the multitude of predictors we could assess, only the preoperative levels of back pain, leg pain and quality of life were influencing their post-operative outcomes. Patient demographic factors, comorbidities or number of treated levels all had insignificant effects. The most successful surgeons were those with very strict selection criteria reflected by high preoperative pain levels and low quality of life. In a situation where alternative treatments like ALIF or even an intensive rehabilitation programme, possibly combined with cognitive behavioural therapy, exist, the question must be asked if a new therapy with still unknown long-term outcomes and risks as well as clinically unproven theoretical advantages like the prevention of adjacent segment disease must not be applied more carefully and selectively. Regulating the application of a therapy with strict inclusion criteria or monitoring its use and related outcomes are viable options. Measuring, e.g. in the framework of a nationwide registry does probably have advantages like quality control and additional evidence generation for these new treatments or their alternatives. In any case, a first step towards improvement of outcome quality could be education about making indications by the best performers. If low outcome quality persists in poorer performing centres, further measures do of course

need to be discussed, but will probably largely depend on frameworks of the respective healthcare systems.

Limitations

We studied three major influential factors (surgeon, implant, procedure), thereby adjusting for those co-variables that were available in the registry. Patients from mandatory and voluntary registers may have different characteristics that cannot completely be controlled with the registry data set. In the current study only age and sex were significantly different, which is, however, rather attributable to the intervention than to the registry setting. Other co-variables with influence on post-operative pain levels and quality of life may also exist. Furthermore, a clear domination of both types of pain may not always be given. Some patients may be treated for high preoperative back pain levels with low or no preoperative leg pain, which may have led to good probabilities for back pain MCRPI but low ones for leg pain. In non-anonymised comparisons, such added complexities must be considered.

Finally, the study was based on 1-year follow-ups only. Although longer follow-ups would be desirable, there is evidence that the results of spinal surgery are often quasi final from the 3-month follow-up onwards [8].

Conclusions

Remarkable variations of pain alleviation by different surgeons were observed. Statistical analysis confirmed selection or indication criteria as at least one of the causes. Although influence of surgeon or implant on pain alleviation after total disc arthroplasty is not significant based on the presented data, it may be clinically relevant. The only significantly influencing co-variate remains the preoperative pain level.

Acknowledgments We thank Daniel Dietrich, PhD, for statistical consulting in all analyses presented in the current article. We are indebted to the SWISSspine and Spine Tango registry groups who made this research possible by populating the database with their valuable and much appreciated entries. The analysed data were recorded by: Bärlocher C ($n = 76$), Sgier F ($n = 65$), Etter C ($n = 41$), Hausmann O ($n = 40$), Schwarzenbach O ($n = 38$), Huber J ($n = 36$), Aebi M ($n = 31$), Heini P ($n = 23$), Berlemann U ($n = 23$), Markwalder T ($n = 19$), Otten P ($n = 17$), Schaeren S ($n = 16$), Maestretti GL ($n = 12$), Schizas C ($n = 12$), Waelchli B ($n = 12$), Porchet F ($n = 10$), Baur M ($n = 9$), Kast E ($n = 9$), Seidel U ($n = 9$), Lutz T ($n = 7$), Grob D ($n = 6$), Jeanneret B ($n = 5$), Kroeber M ($n = 5$), Min K ($n = 5$), Hasdemir M ($n = 4$), Lattig F ($n = 4$), Morard M ($n = 4$), Renella R ($n = 4$), Richter H ($n = 4$), Van Domelen K ($n = 4$), Wernli FO ($n = 4$), Binggeli R ($n = 3$), Stoll TM ($n = 3$), Marchesi D ($n = 3$), Tessitore E ($n = 3$), Vernet O ($n = 3$), Faundez A ($n = 2$), Favre J ($n = 2$), Ramadan A ($n = 2$), Selz T ($n = 2$), Boos N ($n = 1$), Cathrein P ($n = 1$), Forster

T ($n = 1$), Heilbronner R ($n = 1$), Kleinstueck F ($n = 1$), Martinez R ($n = 1$), Rischke B ($n = 1$).

Conflict of interest None of the authors has any potential conflict of interest.

References

- Schlussmann E, Aghayev E, Staub L, Moulin P, Zweig T, Roder C (2010) SWISSspine: the case of a governmentally required HTA-registry for total disc arthroplasty: results of cervical disc prostheses. *Spine (Phila Pa 1976)* 35:E1397–E1405. doi:[10.1097/BRS.0b013e3181e0e871](https://doi.org/10.1097/BRS.0b013e3181e0e871)
- Schlussmann E, Diel P, Aghayev E, Zweig T, Moulin P, Roder C (2009) SWISSspine: a nationwide registry for health technology assessment of lumbar disc prostheses. *Eur Spine J* 18:851–861. doi:[10.1007/s00586-009-0934-8](https://doi.org/10.1007/s00586-009-0934-8)
- Aghayev E, Roder C, Zweig T, Etter C, Schwarzenbach O (2010) Benchmarking in the SWISSspine registry: results of 52 Dynardi lumbar total disc replacements compared with the data pool of 431 other lumbar disc prostheses. *Eur Spine J* 19:2190–2199. doi:[10.1007/s00586-010-1550-3](https://doi.org/10.1007/s00586-010-1550-3)
- Melloh M, Staub L, Aghayev E, Zweig T, Barz T, Theis JC, Chavanne A, Grob D, Aebi M, Roeder C (2008) The international spine registry SPINE TANGO: status quo and first results. *Eur Spine J* 17:1201–1209. doi:[10.1007/s00586-008-0665-2](https://doi.org/10.1007/s00586-008-0665-2)
- Roder C, Staub L, Dietrich D, Zweig T, Melloh M, Aebi M (2009) Benchmarking with Spine Tango: potentials and pitfalls. *Eur Spine J* 18(Suppl 3):305–311. doi:[10.1007/s00586-009-0943-7](https://doi.org/10.1007/s00586-009-0943-7)
- Zweig T, Mannion AF, Grob D, Melloh M, Munting E, Tuschel A, Aebi M, Roder C (2009) How to Tango: a manual for implementing Spine Tango. *Eur Spine J* 18(Suppl 3):312–320. doi:[10.1007/s00586-009-1074-x](https://doi.org/10.1007/s00586-009-1074-x)
- Kessler JT, Melloh M, Zweig T, Aghayev E, Roder C (2011) Development of a documentation instrument for the conservative treatment of spinal disorders in the International Spine Registry, Spine Tango. *Eur Spine J* 20:369–379. doi:[10.1007/s00586-010-1474-y](https://doi.org/10.1007/s00586-010-1474-y)
- Mannion AF, Porchet F, Kleinstueck FS, Lattig F, Jeszenszky D, Bartanusz V, Dvorak J, Grob D (2009) The quality of spine surgery from the patient's perspective. Part 1: the Core Outcome Measures Index in clinical practice. *Eur Spine J* 18(Suppl 3):367–373. doi:[10.1007/s00586-009-0942-8](https://doi.org/10.1007/s00586-009-0942-8)
- Errico TJ (2005) Lumbar disc arthroplasty. *Clin Orthop Relat Res*:106–117. pii:00003086-200506000-00016
- Hagg O, Fritzell P, Nordwall A (2003) The clinical importance of changes in outcome scores after treatment for chronic low back pain. *Eur Spine J* 12:12–20. doi:[10.1007/s00586-002-0464-0](https://doi.org/10.1007/s00586-002-0464-0)
- Roder C, Chavanne A, Mannion AF, Grob D, Aebi M (2005) SSE Spine Tango—content, workflow, set-up. *Eur Spine J* 14:920–924. doi:[10.1007/s00586-005-1023-2](https://doi.org/10.1007/s00586-005-1023-2). www.eurospine.org/SpineTango