**Online supplementary material**

1. **Calculation of risk-adjusted outcome rates:**

To adjust for patient case-mix for each of the outcomes of interest, we performed a logistic regression model adjusting for variables that were demonstrated to be important in a prior study, including age, sex, insurance status, race, American Society of Anesthesiologist score (ASA), smoking status, prior spine surgery, diagnosis (spondylolisthesis, disc herniation, post laminectomy/failed back syndrome, stenosis, pseudarthrosis, radiculopathy), opiate use, asthma, and baseline ODI scores.1

The probability of improvement beyond MID or reaching minimal disability level for each patient was then estimated and summed for each hospital or surgeon to arrive at an “expected number of events”. A hospital-or-surgeon specific observed/expected (O/E) ratio was then calculated by dividing the observed number of events for each hospital/surgeon by the calculated expected number. Finally, this O/E ratio was multiplied by the overall statewide average event rate to yield a risk-adjusted event rate for each hospital or surgeon.

We reported the intraclass correlation coefficients (ICC) before and after risk adjustment at the hospital and surgeon level, which describe the proportion of total variability accounted for by between-hospital variance and between-surgeon variance.

1. **Detailed information on reliability adjustment**

Reliability adjustment was used to account for the proportion of observed variation in outcomes that could be due to chance alone.2 This technique weighs the observed rate based on the measure’s reliability, which reflects the extent to which the measure is reproducible rather than random, and attributes the remaining weight to the overall mean. The reliability-adjusted rates shrink towards the overall mean and the shrinkage is greater for hospitals or surgeons with low volume. However, there are concerns that this approach tends to diminish observed differences among providers, especially among low-volume providers, and concomitantly penalizes or rewards high-volume providers.3

To adjust the outcomes for reliability, we used the empirical Bayes’ techniques. First we performed a logistic regression model for each outcome using the same patient characteristics in our risk-adjustment model to predict each patient’s risk score (i.e. log(odds) of an outcome). Then, for each of the outcomes of interest, we created a 2-level hierarchical logistic model (meglm command in STATA) where this patient risk score was entered as the fixed effect (first level) and the hospital or surgeon identifier was entered as the random intercept (the second level). The random effect from this model (in log(odds)) was then added to the average patient risk score. The inverse logit of this sum is the risk-and-reliability-adjusted rate. The details of this procedure are recorded elsewhere.4

**eFigure 1a:** Unadjusted proportion of patients experiencing a MID improvement in ODI at 12 months (ODI decreased by ≥15) by hospital (ICC = 0.042). Horizontal line indicates overall unadjusted improvement rate. Hospitals with <10 cases are not shown.

**eFigure 1b:** Figure 2a: Unadjusted proportion of patients who reached minimal disability at 12 months (ODI<22) by hospital (ICC = 0.010). Horizontal line indicates overall unadjusted rate. Hospitals with <10 cases are not shown.

**eFigure 2a:** Unadjusted proportion of patients experiencing a MID improvement in ODI at 12 months (ODI decreased by ≥15) by surgeon (ICC = 0.065). Horizontal line indicates overall unadjusted rate. Surgeons with <10 cases are not shown.

**eFigure 2b:** Unadjusted proportion of patients who reached minimal disability at 12 months following lumbar fusion surgery (ODI<22) by surgeon (ICC=0.044). Horizontal line indicates overall unadjusted rate. Surgeons with <10 cases are not shown.

**References**

1. Khor S, Lavallee D, Cizik AM, et al. Development and validation of a prediction model for pain and functional outcomes after lumbar spine surgery. *JAMA Surg*. 2018. doi: 10.1001/jamasurg.2018.0072 [doi].

2. MacKenzie TA, Grunkemeier GL, Grunwald GK, et al. A primer on using shrinkage to compare in-hospital mortality between centers. *Ann Thorac Surg*. 2015;99(3):757-761. doi: 10.1016/j.athoracsur.2014.11.039.

3. Weintraub WS, Garratt KN. Challenges in risk adjustment for hospital and provider outcomes assessment. *Circulation*. 2017;135(4):317-319. doi: 10.1161/CIRCULATIONAHA.116.025653.

4. Dimick JB, Ghaferi AA, Osborne NH, Ko CY, Hall BL. Reliability adjustment for reporting hospital outcomes with surgery. *Ann Surg*. 2012;255(4):703-707. doi: 10.1097/SLA.0b013e31824b46ff.