




BMJ Open Mortality as an indicator of quality of neurosurgical care in England: a retrospective cohort study

Adam J Wahba ^{1,2} David A Cromwell,^{1,3} Peter J Hutchinson ^{4,5}
Ryan K Mathew ^{2,6} Nick Phillips^{6,7}

To cite: Wahba AJ, Cromwell DA, Hutchinson PJ, *et al*. Mortality as an indicator of quality of neurosurgical care in England: a retrospective cohort study. *BMJ Open* 2022;**12**:e067409. doi:10.1136/bmjopen-2022-067409

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-067409>).

Received 12 August 2022
Accepted 07 October 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Clinical Effectiveness Unit, Royal College of Surgeons, London, UK

²Leeds Institute of Medical Research, School of Medicine, University of Leeds, Leeds, UK

³Health Services Research & Policy, London School of Hygiene & Tropical Medicine, London, UK

⁴Academic Neurosurgery, University of Cambridge, Cambridge, UK

⁵Clinical Research, Royal College of Surgeons, London, UK

⁶Department of Neurosurgery, Leeds Teaching Hospitals NHS Trust, Leeds, UK

⁷Clinical Lead for Cranial Neurosurgery, Getting It Right First Time (GIRFT), London, UK

Correspondence to

Mr Adam J Wahba;
adam.wahba@nhs.net

ABSTRACT

Objectives Postoperative mortality is a widely used quality indicator, but it may be unreliable when procedure numbers and/or mortality rates are low, due to insufficient statistical power. The objective was to investigate the statistical validity of postoperative 30-day mortality as a quality metric for neurosurgical practice across healthcare providers.

Design Retrospective cohort study.

Setting Hospital Episode Statistics data from all neurosurgical units in England.

Participants Patients who underwent neurosurgical procedures between April 2013 and March 2018. Procedures were grouped using the National Neurosurgical Audit Programme classification.

Outcomes measured National 30-day postoperative mortality rates were calculated for elective and non-elective neurosurgical procedural groups. The study estimated the proportion of neurosurgeons and NHS trusts in England that performed sufficient procedures in 3-year and 5-year periods to detect unusual performance (defined as double the national rate of mortality). The actual difference in mortality rates that could be reliably detected based on procedure volumes of neurosurgeons and units over a 5-year period was modelled.

Results The 30-day mortality rates for all elective and non-elective procedures were 0.4% and 6.1%, respectively. Only one neurosurgeon in England achieved the minimum sample size (n=2402) of elective cases in 5 years needed to detect if their mortality rate was double the national average. All neurosurgical units achieved the minimum sample sizes for both elective (n=2402) and non-elective (n=149) procedures. In several neurosurgical subspecialties, approximately 80% of units (or more) achieved the minimum sample sizes needed to detect if their mortality rate was double the national rate, including elective neuro-oncology (baseline mortality rate=2.3%), non-elective neuro-oncology (rate=5.7%), neurovascular (rate=6.7%) and trauma (rate=11%).

Conclusion Postoperative mortality lacks statistical power as a measure of individual neurosurgeon performance. Neurosurgical units in England performed sufficient procedure numbers overall and in several subspecialty areas to support the use of mortality as a quality indicator.

INTRODUCTION

Early postoperative mortality is a commonly used metric in research and quality assurance

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ The study evaluated the statistical validity of 30-day postoperative mortality as a metric for evaluating the quality of neurosurgical care.
- ⇒ Population-based cohorts were derived from hospital administrative data that included all patients who underwent neurosurgical procedures in the National Health Service in England.
- ⇒ The study used power and sample size analyses to compute the minimum sample sizes needed to detect a mortality rate that was double the national average.
- ⇒ A limitation of the analysis was that it focused solely on statistical validity which is just one of several aspects of the validation of quality metrics.
- ⇒ Routinely collected administrative data may be subject to errors in the coding and procedures and diagnoses, which could affect the classification of neurosurgical procedures.

programmes. The investigation of national, institutional and surgeon-level outcomes is seen as one way of measuring quality of care and reducing the frequency of adverse events.¹ However, several reports have shown that postoperative mortality may not be a statistically valid measure when mortality is an infrequent outcome, and/or procedure numbers are low.² Low procedure numbers may lead to complacency about quality of care because poor performance may not be detectable.³

National neurosurgical quality improvement programmes in both the UK and USA have been criticised as having insufficient statistical power to report reliable comparative outcomes.⁴ The UK National Neurosurgical Audit Programme (NNAP) was estimated to need sample sizes that would far exceed the volume of procedures that a neurosurgeon would perform in a typical audit cycle to detect excess mortality which might be linked to poor performance.⁴ It is therefore necessary for the statistical validity of neurosurgical

Table 1 The number of elective and non-elective neurosurgical procedures performed between April 2013 and March 2018 in England

Procedure category	No of procedures		Total (n)
	Elective (n, %)	Non-elective (n, %)	
I. All procedures and major procedures			
All procedures	190 127 (67.6)	91 042 (32.4)	281 169
Major procedures	150 575 (64.8)	81 970 (35.2)	232 545
II. Cranial and spinal procedures			
Cranial surgery	64 953 (51.8)	60 437 (48.2)	125 390
Spinal surgery	85 622 (79.9)	21 534 (20.1)	107 156
III. Subspecialty procedures			
Trauma	—	13 638 (100)	13 638
General neurosurgery	5202 (28.8)	12 847 (71.2)	18 049
Functional	22 166 (95.0)	1162 (5.0)	23 328
Neuro-oncology	17 520 (68.5)	8041 (31.5)	25 561
All CSF procedures	9711 (44.0)	12 365 (56.0)	22 076
CSF shunt procedures	6280 (47.1)	7061 (52.9)	13 341
Skull base	12 787 (85.9)	2098 (14.1)	14 885
Neurovascular	6661 (38.5)	10 650 (61.5)	17 311
Simple spine	64 644 (81.4)	14 786 (18.6)	79 430
Complex spine	8382 (62.2)	5099 (37.8)	13 481
Intradural spine	3205 (71.7)	1262 (28.3)	4467
CSF, cerebrospinal fluid.			

postoperative mortality at individual practitioner and institutional level to be investigated in order to provide confidence in quality improvement programmes that use it as a quality indicator. Where statistical power is inadequate, mortality rates may not be a suitable measure of quality and the focus should shift to developing alternative indicators.

We hypothesised that postoperative 30-day mortality has statistical validity as a quality metric for neurosurgery. We sought to explore how it can be used to evaluate neurosurgical practice. Specific objectives were to calculate national mortality rates for neurosurgical procedures, calculate procedure volumes and estimate if procedure volumes are sufficient to allow detection of statistical outliers at both the individual and institutional level.

METHODS

Data and study design

This was a retrospective cohort study of routinely collected hospital data using a 5-year extract of Hospital Episode Statistics (HES) data from 1 April 2013 to 31 March 2018. HES is the hospital administrative database for the National Health Service (NHS) in England. It contains information on surgical procedures, diagnoses and administrative data. Diagnoses are coded using the International Classification of Diseases-10. Procedures are coded using the classification from the UK Office of

Population Censuses and Surveys (OPCS, version 4). The extract contained summary patient records where neurosurgery was the main specialty, and a primary neurosurgical procedure was performed in one of the 24 NHS neurosurgical units in England. The study participants were patients who underwent a neurosurgical procedure identified using the NNAP Coding Framework, a database of approximately 870 OPCS codes (online supplemental table S1). The study pooled procedures into groups within three tiers (online supplemental table S2). Procedures were allocated to the appropriate group (in tier III) using the primary diagnosis as a filter where necessary. The UK General Medical Council specialist register for neurosurgery was used to identify Consultant Neurosurgeons and calculate surgeon-level procedure volumes.⁵ The primary outcome measure was postoperative all-cause 30-day mortality. Information on the date of death was obtained from the Civil Registration of Mortality data.⁶

Statistical analysis

The analysis proceeded in a series of steps, with elective and non-elective procedures considered separately. First, national 30-day mortality rates were calculated for each group. Second, 3-year and 5-year total procedure volumes, and average annual volumes per neurosurgeon and per unit were estimated for each group.

Third, we calculated the minimum sample size needed to detect if a provider's (neurosurgeon/neurosurgery unit)

mortality rate was double the national average, and the proportion of providers achieving the minimum sample sizes in three year and five year periods was estimated. The study used power and sample size analyses to compute the minimum sample sizes, with calculations powered at 80%. That is, the chance of a type II error (false negative) in detecting a provider with double the baseline mortality rate was 20%. A significance level of 0.05 was used. In practice, a doubling in mortality would represent a significant difference in the number of patient deaths and is a reasonable threshold for poor performance. Quality assurance programmes use statistical methods to evaluate outcomes such as funnel plots or control charts.^{7,8} Funnel plots use control limits to define the acceptable level of variation from the average outcome (usually two or three SD from the mean), and providers that breach of these limits are considered outliers. A doubling in mortality would result in most providers breaching the control limits, except those with the smallest procedure volumes. This approach is consistent with previous studies in this area.^{2,3,9}

Finally, the study modelled the relative difference in mortality from the national average that could be reliably detected for different sample sizes. In addition, the difference in mortality that could be detected for the providers that perform the lowest volumes in five years was estimated. The 10th percentile of provider volume was used—rather than the absolute lowest—to avoid basing the calculation on an anomaly, such as a hospital that does not routinely perform the procedures.

Data analysis was performed using Stata, V.15 (StataCorp). The study followed guidance suggested by the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement.¹⁰

Patient and public involvement

None.

RESULTS

From April 2013 to March 2018, there were 281 169 neurosurgical procedures performed in the 24 neurosurgical

Table 2 The number of providers that perform enough elective neurosurgical procedures over 3-year and 5-year periods to detect statistical outliers if the 30-day mortality rate is double the national average

Elective neurosurgery	National average mortality rate for elective neurosurgery (%)	Average annual no of procedures per surgeon/hospital (n)	Minimum sample size (80% power) (n)	Surgeons/hospitals achieving or exceeding minimum sample size	
				3 years (n, %)	5 years (n, %)
Surgeons					
I. All procedures and major procedures					
All procedures	0.4	96	2402	0/398 (0)	1/407 (0.2)
Major procedures	0.5	81	1989	0/390 (0)	1/399 (0.3)
Hospitals					
I. All procedures and major procedures					
All procedures	0.4	1481	2402	21/24 (87.5)	24/24 (100)
Major procedures	0.5	1192	1989	22/24 (91.7)	23/24 (95.8)
II. Cranial and spinal procedures					
Cranial surgery	1.0	452	969	17/24 (70.8)	22/24 (91.7)
Spinal surgery	0.1	663	9423	0/24 (0)	0/24 (0)
III. Subspecialty procedures					
General neurosurgery	0.7	39	1345	0/24 (0)	0/24 (0)
Functional	0.2	138	4567	0/24 (0)	0/24 (0)
Neuro-oncology	2.3	145	410	14/24 (58.3)	19/24 (79.2)
All CSF procedures	0.5	60	2176	0/24 (0)	0/24 (0)
CSF shunt procedures	0.4	40	2294	0/24 (0)	0/24 (0)
Skull base	0.7	101	1445	0/24 (0)	0/24 (0)
Neurovascular	0.7	54	1488	0/24 (0)	0/24 (0)
Simple spine	0.1	524	12 559	0/24 (0)	0/24 (0)
Complex spine	0.2	59	5925	0/24 (0)	0/24 (0)
Intradural spine	0.3	26	3519	0/24 (0)	0/24 (0)
CSF, cerebrospinal fluid.					

Table 3 The number of providers that perform enough non-elective neurosurgical procedures over 3-year and 5-year periods to detect statistical outliers if the 30-day mortality rate is double the national average

Non-elective neurosurgery	National average mortality rate for non-elective neurosurgery (%)	Average annual no of procedures per surgeon/hospital (n)	Minimum sample size (80% power) (n)	Surgeons/hospitals achieving or exceeding minimum sample size	
				3 years (n, %)	5 years (n, %)
Surgeons					
I. All procedures and major procedures					
All procedures	6.1	51	149	149/353 (42.2)	234/370 (63.2)
Major procedures	6.6	47	137	143/349 (41.0)	229/364 (62.9)
Hospitals					
I. All procedures and major procedures					
All procedures	6.1	770	149	24/24 (100)	24/24 (100)
Major procedures	6.6	702	137	24/24 (100)	24/24 (100)
II. Cranial and spinal procedures					
Cranial surgery	8.4	495	105	24/24 (100)	24/24 (100)
Spinal surgery	1.6	162	617	9/24 (37.5)	19/24 (79.2)
III. Subspecialty procedures					
Trauma	11.0	112	77	23/24 (95.8)	24/24 (100)
General neurosurgery	7.6	110	118	24/24 (100)	24/24 (100)
Functional	3.4	7	282	0/24 (0)	0/24 (0)
Neuro-oncology	5.7	66	160	16/24 (66.7)	23/24 (95.8)
All CSF procedures	10.7	109	79	24/24 (100)	24/24 (100)
CSF shunt procedures	3.0	63	317	0/24 (0)	12/24 (50.0)
Skull base	2.3	15	412	0/24 (0)	0/24 (0)
Neurovascular	6.7	91	136	20/24 (83.3)	24/24 (100)
Simple spine	1.2	116	811	0/24 (0)	8/24 (33.3)
Complex spine	2.5	43	388	0/24 (0)	1/24 (4.2)
Intradural spine	2.6	10	366	0/24 (0)	0/24 (0)
CSF, cerebrospinal fluid.					

units in England. Of these, 232 545 (82.7%) were major neurosurgical procedures (defined in online supplemental table S2). The total numbers of procedures in each group are shown in [table 1](#).

Thirty-day mortality rates for elective neurosurgical procedures ranged from 0% to 0.7%, except for neuro-oncology at 2.3%. Non-elective mortality rates ranged from 0% to 11.0%. The average annual volume of elective and non-elective procedures per neurosurgeon were 96 and 51, respectively, and per hospital were 1481 and 770, respectively. The highest average annual volumes for elective neurosurgery per hospital were in simple spine surgery (median volume, n=524), neuro-oncology (n=145) and functional neurosurgery (n=138). The highest volumes for non-elective neurosurgery were in simple spine surgery (n=116), trauma (n=112) and general neurosurgery (n=110) ([tables 2 and 3](#)).

No individual Consultant Neurosurgeon in England achieved the minimum sample size needed for all procedures (n=2402) or major procedures (n=1989) to detect

a doubling of the baseline elective 30-day mortality rate in 3 years, and only one neurosurgeon achieved this threshold in 5 years. Conversely, in non-elective neurosurgery, 63% of neurosurgeons achieved the minimum sample size for all procedures (n=149) and major procedures (n=137), which reflected the higher average mortality rates associated with these procedures.

Extending the period of evaluation from 3 to 5 years increased the proportion of neurosurgical units achieving the minimum sample size in several procedure groups. The distribution plots in [figure 1](#) demonstrate where unit-level volumes met the minimum sample size threshold. The hierarchy diagrams in [figure 2](#) explore the effect of pooling procedures; pooling increases the sample size but this also increases the heterogeneity of the procedures within the groups as indicated by the different post-operative mortality rates within the lower tiers.

The difference in 30-day mortality rates that could be reliably detected for the neurosurgical units that performed the lowest volume (10th percentile) of procedures are

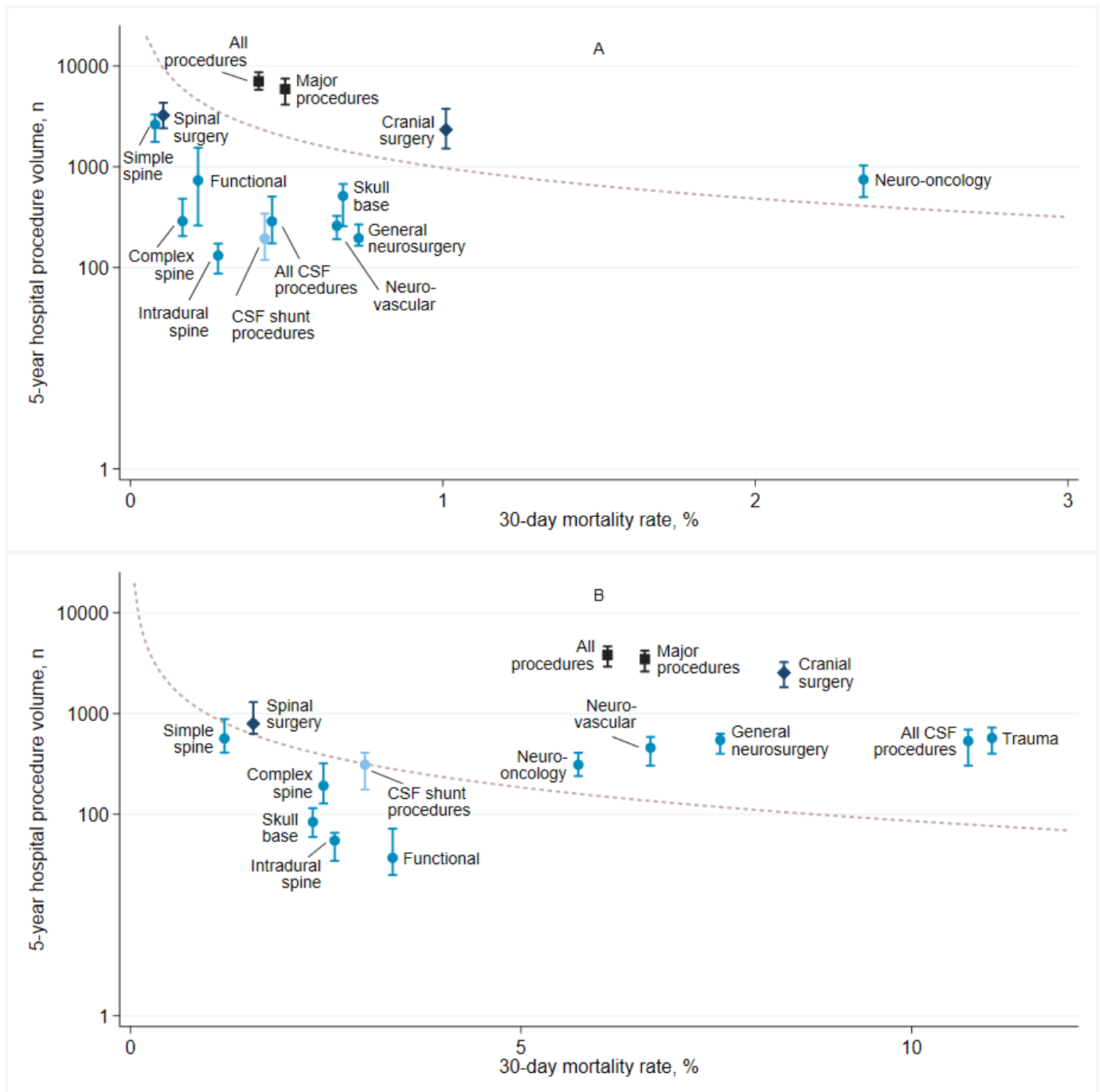


Figure 1 Distribution plot showing the 5-year hospital procedure volume for elective (A) and non-elective (B) neurosurgical procedures and the minimum sample threshold needed to detect a doubling of the national 30-day mortality rate. The 'I-bars' show the 75th percentile (top cap), median (marker) and 25th percentile (bottom cap) of hospital procedure volume. The dashed line shows the threshold for minimum procedure volume; the area above this indicates where Hospital procedure volumes exceed the threshold. CSF - cerebrospinal fluid.

shown in tables 4 and 5. The ratios ranged from 1.8 to 11.8 times the baseline rate for elective neurosurgery and 1.2–6.8 times the rate for non-elective neurosurgery. Importantly, in several categories the lowest volume units performed sufficient numbers of procedures to detect smaller deviations (less than a doubling) in mortality rates. This occurred more frequently in non-elective neurosurgery, where baseline mortality rates are much higher and the minimum sample sizes are consequently smaller.

Figure 3 demonstrates the sample sizes required to detect a range of relative increases in mortality compared with the national average for all procedures.

DISCUSSION

This study aimed to evaluate the statistical validity of 30-day postoperative mortality rates as a quality indicator for neurosurgery, at both individual practitioner and

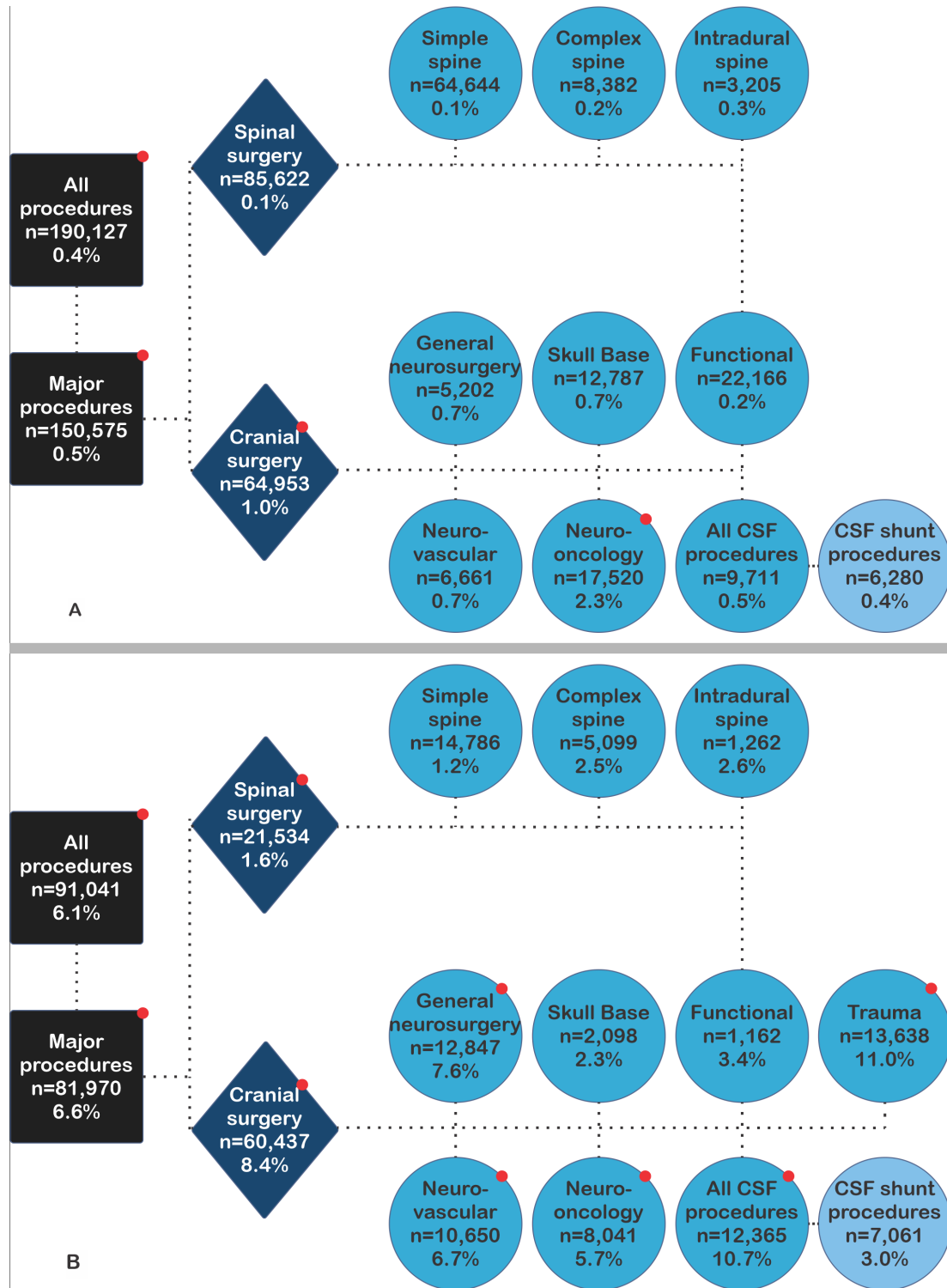


Figure 2 Hierarchy diagrams demonstrating how elective (A) and non-elective (B) neurosurgical procedures were pooled for analysis in three tiers. The volume of procedures included in the analysis and the 30-day mortality rates are shown for each group. The red marker shows the groups for which the majority of neurosurgical units performed a sufficient volume of procedures to surpass the minimum sample size threshold to detect a doubling in the baseline mortality rate. Increasing volume by pooling procedures provides greater statistical power, but it also increases the heterogeneity of the group. CSF - cerebrospinal fluid. This figure was produced using CoreIDRAW 2020 (Corel, Ottawa, Canada).

institutional level, and to explore the use of this metric in evaluating neurosurgical practice. Early postoperative mortality is a metric commonly used by quality assurance programmes, including the NNAP in the UK, the National

Surgical Quality Improvement Programme (NSQIP) and Quality Outcomes Database (QOD) in the USA,^{5 11 12} the recently established Mayo Clinic neurosurgical registry and the Japan National Database.^{13 14}

Table 4 Increases in mortality rates compared with the national average that could be reliably detected for providers that perform the lowest volume of elective neurosurgical procedures in 5 years

Elective neurosurgery	National average mortality rate (%)	Change in mortality rate that could be detected for surgeons/hospitals that perform lowest volumes of procedures	
		Mortality rate (%)	Ratio to national average
Surgeons			
I. All procedures and major procedures			
All procedures	0.4	3.8	9.3 times
Major procedures	0.5	4.2	8.5 times
Hospitals			
I. All procedures and major procedures			
All procedures	0.4	0.7	1.8 times
Major procedures	0.5	0.9	1.8 times
II. Cranial and spinal procedures			
Cranial surgery	1.0	2.1	2.0 times
Spinal surgery	0.1	0.4	3.7 times
III. Subspecialty procedures			
General neurosurgery	0.7	3.8	5.2 times
Functional	0.2	1.8	8.5 times
Neuro-oncology	2.3	5.4	2.3 times
All CSF procedures	0.5	3.3	7.2 times
CSF shunt procedures	0.4	3.9	9.0 times
Skull base	0.7	3.1	4.6 times
Neurovascular	0.7	4.2	6.4 times
Simple spine	0.1	0.5	5.7 times
Complex spine	0.2	1.6	9.5 times
Intradural spine	0.3	3.3	11.8 times

CSF, cerebrospinal fluid.

Postoperative mortality lacks statistical power as a measure of performance for individual neurosurgeons, and for neurosurgical units in many areas of practice. Procedure volumes are generally low, and mortality is generally an uncommon outcome. After elective neurosurgery, the mortality rate is 0.4% (all procedures) and it is <1% in most areas of subspecialty practice (except neuro-oncology at 2.3%). Mortality is higher after non-elective neurosurgery at 6.1% and lies between 0% and 11.0% in subspecialty practice.

Surgeon-level outcomes

The UK government made publication of surgical outcomes mandatory in 2013 and the Society of British Neurological Surgeons (SBNS) subsequently established the NNAP to evaluate surgeon-level and hospital-level 30-day mortality rates for elective neurosurgery.⁵ The mandatory publication of surgeons' outcomes in England was justified as a mechanism to identify poor performance, provide transparency and provide information for patients and professionals to support patient choice.¹⁵ However, there are reasons to be cautious about the reliability, fairness and unintended consequences

of publishing surgeons' outcomes. Adverse outcomes after neurosurgery are unpreventable in some cases, and where medical errors do occur the underlying causes may be broader system-level problems or a series of events, as well as errors by individual clinicians.¹⁶ A narrow focus on this metric does not account for the role of surgeons in the wider healthcare team.¹⁷ Surgeons that are incorrectly identified as having poor performance could be stigmatised following inquiries into their outcomes.¹⁸ Risk-averse behaviour among surgeons could lead to the avoidance of complex or high-risk cases and could affect the quality and quantity of training opportunities.^{19 20} Patients may not understand how to interpret mortality data, particularly for operations where outcomes are intrinsically poor.¹⁸

In addition to the potential issues highlighted above is the lack of statistical validity of using surgeon-level mortality rates to detect poor performance. Neurosurgeons in England do not perform sufficient volumes of procedures to detect outlier mortality rates for elective neurosurgical procedures. For surgeons with the lowest procedure volumes, increases in mortality of 9.3 times

Table 5 Increases in mortality rates compared with the national average that could be reliably detected for providers that perform the lowest volume of non-elective neurosurgical procedures in 5 years

Non-elective neurosurgery	National average mortality rate (%)	Change in mortality rate that could be detected for surgeons/hospitals that perform lowest volumes of procedures	
		Mortality rate (%)	Ratio to national average
Surgeons			
I. All procedures and major procedures			
All procedures	6.1	19.7	3.2 times
Major procedures	6.6	20.3	3.1 times
Hospitals			
I. All procedures and major procedures			
All procedures	6.1	7.5	1.2 times
Major procedures	6.6	8.1	1.2 times
II. Cranial and spinal procedures			
Cranial surgery	8.4	10.5	1.3 times
Spinal surgery	1.6	3.7	2.3 times
III. Subspecialty procedures			
Trauma	11.0	17.5	1.6 times
General neurosurgery	7.6	12.1	1.6 times
Functional	3.4	22.7	6.8 times
Neuro-oncology	5.7	11.1	1.9 times
All CSF procedures	10.7	16.9	1.6 times
CSF shunt procedures	3.0	8.0	2.7 times
Skull base	2.3	11.7	5.0 times
Neurovascular	6.7	12.1	1.8 times
Simple spine	1.2	3.5	2.9 times
Complex spine	2.5	10.7	4.3 times
Intradural spine	2.6	15.3	5.9 times

CSF, cerebrospinal fluid.

the national average (3.8%) for all procedures and 8.5 times (4.2%) for major procedures would be detectable. With increasing volume, smaller deviations would be detectable, but the magnitude of the differences goes far beyond what could reasonably be considered poor performance.

Postoperative mortality rates are much higher in non-elective neurosurgery and the minimum sample sizes are therefore lower, such that 63% of surgeons reached the threshold in five years in the major and all procedure groups. While comparisons based on non-elective activity would be more statistically robust, it is doubtful that this offers a meaningful way of identifying poorly performing surgeons; the problems pertaining to accuracy of case attribution and the complexities of emergency care are summarised in online supplemental table S3.

Hospital-level outcomes and subspecialty practice

Hospital-level outcomes should provide a more robust means of evaluating quality of care. All neurosurgical units in England performed enough elective procedures

in 5 years to detect statistical outliers. For the lowest volume units, an increase in mortality of 1.8 times the national average could be detected.

Assessing the performance of hospitals based on emergency neurosurgical care can be a useful way of demonstrating variation in quality and eliciting examples of both good and poor services.^{21–23} All neurosurgical units performed sufficient volumes—in the all, major, cranial and spinal procedure groups—to detect statistical outliers. For most neurosurgical subspecialties, procedure volumes were low and the minimum sample sizes were not achieved in many areas. However, in areas of practice with higher mortality rates—particularly in non-elective practice—many units achieved or exceeded the minimum sample size. This presents the opportunity to use mortality as a metric for neurosurgical subspecialties. Several examples are discussed here.

In elective neurosurgery, neuro-oncology was a subspecialty where 19 of 24 units crossed the sample size threshold. One study has previously explored 30-day mortality rates

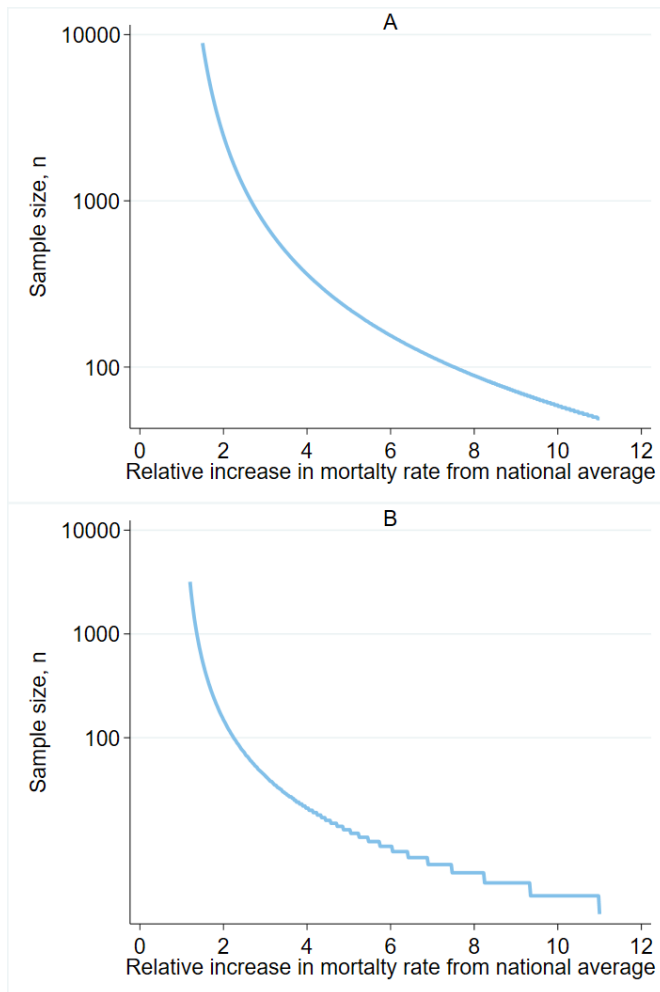


Figure 3 Modelling of minimum sample sizes required to detect relative increases in mortality rates from the national average for (A) all elective neurosurgical procedures (baseline mortality of 0.4%) and (B) all non-elective neurosurgical procedures (baseline mortality of 6.1%). The plots demonstrate the relative increase in mortality that could be reliably detected based on a minimum number of procedures performed by a provider in an audit period.

after resection of brain tumours in England, reporting a median volume of 273 cases per unit (range 29–667).²⁴ It did not observe any instances of higher-than-expected mortality. However, given the minimum sample sizes in our study (410 for elective and 160 for non-elective) it is probable the sample size in some units was too small to detect outliers. Close attention should be paid to the length of the evaluation period and unit-level sample sizes when mortality rates are used as a quality indicator for neuro-oncology.

Trauma neurosurgery had the highest mortality rate at 11%. Mortality is an important outcome measure for traumatic brain injury and data from the national Trauma Audit and Research Network in England and Wales has been used to assess outcomes in over 15 000 patients with traumatic brain injury.²⁵ The study reported several statistical outliers, with both better and worse than expected 30-day mortality rates. In our study, all units exceeded the

minimum caseload threshold (based on 13 638 patients overall), which supports the use of mortality as a quality metric for trauma.

A high mortality rate of 10.7% was observed in non-elective cerebrospinal fluid (CSF) diversion procedures but this falls to 3.0% when only CSF shunt surgery was analysed, with a difference in average annual hospital procedure volume of 42% between the two categories (109 vs 63). Additional procedures in the CSF diversion group include external ventricular drain (EVD) insertion and endoscopic third ventriculostomy; the difference in mortality rates suggests a very high mortality rate in these additional procedures. There are a wide variety of indications for EVD insertion and differing practices in the management of acute hydrocephalus.²⁶ Variation in mortality rates for CSF diversion surgery could be related to differences in procedural casemix, patient selection and local management practices—as well as poor quality care. Comparative outcomes would need to be interpreted in the context of these differences.

Improving statistical power

Several strategies can be employed to improve statistical power, but each introduces its own problems; these are summarised in online supplemental table S4. A key strategy is the pooling of procedures to increase sample size; this increases heterogeneity, which can complicate interpretation of the results and it may introduce problems in relation to risk-adjustment and fairness. Neurosurgical outcomes derived from national data can be used to support quality improvement, but care should be taken to avoid institutional stigma due to unfair or inaccurate comparisons.^{27 28} Random variation in mortality rates between providers is to be expected; appropriate reporting methods, such as funnel plots, should be used to determine if that variation is excessive.⁷ Mortality is undoubtedly a serious adverse outcome but there are a range of alternative quality metrics that are relevant to neurosurgeons and patients alike.¹² The evaluation of mortality rates may have utility in assuring a minimum level of surgical safety, rather than measuring quality, particularly where rates are intrinsically low and variation is difficult to interpret.

The data in this study are from the period prior to the COVID-19 pandemic; the reduction in surgical activity will exacerbate the challenges relating to statistical power.²⁹ Compromises may be necessary to overcome this, such as having extended audit periods until surgical throughput normalises.

Limitations

This study used routinely collected administrative data which is subject to errors in the coding of clinical information. Errors could have affected both the classification of neurosurgical procedures and whether all neurosurgical procedures are captured in HES. This study has focused on statistical power, but other aspects to consider when evaluating quality metrics include risk adjustment,

relevance of the metric to patients and clinicians, and the extent to which the risk of mortality is modifiable. Outcomes must be risk adjusted to improve the reliability of comparative evaluations, but risk-adjustment models for neurosurgery using hospital administrative data alone may not be adequate.³⁰ Procedures were classified by type in this study but the complexity of cases (either clinically or surgically) has not been explored. Factors relating to case complexity could vary between units and ought to be considered in any evaluations of quality, possibly by risk-adjustment or stratification of cases. Our study did not explore if this was possible using these data.

CONCLUSIONS

Neurosurgical 30-day postoperative mortality, when applied at the appropriate provider level (institutional and not individual surgeon), is a statistically valid metric for comparing performance in several areas of neurosurgical practice. The UK NNAP hospital-level outlier programme has the statistical power to detect poor performance in elective and non-elective neurosurgical procedures. Mortality is a valid metric for more focused areas of neurosurgical practice, including elective neuro-oncology surgery and non-elective neuro-oncology surgery, neurovascular surgery, general neurosurgery, trauma neurosurgery and CSF diversion procedures. However, the NNAP surgeon-level outlier programme lacks the statistical power to detect poor performance, due to low procedure volume and the low mortality rates. More research is needed to develop effective strategies for evaluating quality in the context of low procedure volumes.

Twitter Adam J Wahba @adamwahba and Ryan K Mathew @LeedsNeuro

Acknowledgements AJW is supported by an RCS Research Fellowship. The fellowship is jointly based within the Society of British Neurological Surgeons NNAP and the Clinical Effectiveness Unit at the Royal College of Surgeons of England. PJH is supported by the National Institute for Health Research (Senior Investigator Award, Cambridge Biomedical Research Centre, Brain Injury MedTech Co-operative, Global Neurotrauma Research Group) and the Royal College of Surgeons of England. RKM is supported by Yorkshire's Brain Tumour Charity and Candlelighters.

Contributors AJW: study conception, design, data acquisition, data analysis, data interpretation, writing—draft preparation and revision. DAC: study conception, design, data acquisition, data analysis, data interpretation, writing—revision. PJH: data interpretation, data analysis, writing—revision. RKM: data interpretation, data analysis, writing—revision. NP: study conception, design, data interpretation, data analysis, writing—revision. All authors read and approved the final manuscript. Guarantor: AJW accepts full responsibility for the finished work and the conduct of the study, had access to the data, and controlled the decision to publish.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Ethics approval The study is exempt from UK National Research Ethics Service (NRES) approval because it involved the analysis of routinely collected, anonymised data. HES data were made available by NHS Digital (Copyright 2018, Reused with the permission of NHS Digital. All rights reserved). Approvals for the use of anonymised HES data were obtained as part of the standard NHS Digital data access process. The data are routinely collected by the National Health Service (NHS) and anonymised at source. The data are obtained as part of the standard NHS Digital data access process. NHS Digital has a legal

obligation to collect these data and it does not require the consent of individual patients. Please see here: <https://digital.nhs.uk/about-nhs-digital/our-work/keeping-patient-data-safe/gdpr/gdpr-register/hospital-episode-statistics-gdpr/hospital-episode-statistics-hes-gdpr-information>.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement No data are available.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

Adam J Wahba <http://orcid.org/0000-0001-5746-0129>

Peter J Hutchinson <http://orcid.org/0000-0002-2796-1835>

Ryan K Mathew <http://orcid.org/0000-0002-2609-9876>

REFERENCES

- Karhadé AV, Larsen AMG, Cote DJ, *et al*. National databases for neurosurgical outcomes research: options, strengths, and limitations. *Neurosurgery* 2018;83:333–44.
- Dimick JB, Welch HG, Birkmeyer JD. Surgical mortality as an indicator of hospital quality. *JAMA* 2004;292:847.
- Walker K, Neuburger J, Groene O, *et al*. Public reporting of surgeon outcomes: low numbers of procedures lead to false complacency. *The Lancet* 2013;382:1674–7.
- Reponen E, Tuominen H, Korja M. Quality of British and American nationwide quality of care and patient safety benchmarking programs: case neurosurgery. *Neurosurgery* 2019;85:500–7.
- Society of British Neurological Surgeons (SBNS). The neurosurgical national audit programme (NNAP), 2021. Available: <https://www.nnap.org.uk/> [Accessed 17 Jun 2021].
- NHS Digital. Linked HES-ONS mortality data, 2022. Available: <https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/linked-hes-ONS-mortality-data> [Accessed 18 Feb 2022].
- Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med* 2005;24:1185–202.
- Neuburger J, Walker K, Sherlaw-Johnson C, *et al*. Comparison of control charts for monitoring clinical performance using binary data. *BMJ Qual Saf* 2017;26:919–28.
- Harrison EM, Drake TM, O'Neill S, *et al*. Individual surgeon mortality rates: can outliers be detected? A national utility analysis. *BMJ Open* 2016;6:e012471–6.
- Von Elm E, Altman DG, Egger M, *et al*. The strengthening of reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. *Int J Surg* 2014;12:1495–9.
- Rock AK, Opalak CF, Workman KG, *et al*. Safety outcomes following spine and cranial neurosurgery: evidence from the National surgical quality improvement program. *J Neurosurg Anesthesiol* 2018;30:328–36.
- Parker SL, McGirt MJ, Bekelis K, *et al*. The National neurosurgery quality and outcomes database qualified clinical data registry: 2015 measure specifications and rationale. *Neurosurg Focus* 2015;39:E4.
- Iihara K, Tominaga T, Saito N, *et al*. The Japan neurosurgical database: overview and results of the first-year survey. *Neurol Med Chir* 2020;60:165–90.
- Bydon M, Goyal A, Biedermann A, *et al*. Building and implementing an institutional registry for a data-driven national neurosurgical practice: experience from a multisite medical center. *Neurosurg Focus* 2021;51:E9.
- Alderson D, Cromwell D. Publication of surgeon-specific outcomes. *Br J Surg* 2014;101:1335–7.

- 16 Rolston JD, Bernstein M. Errors in neurosurgery. *Neurosurg Clin N Am* 2015;26:149–55.
- 17 Khajuria A. Public reporting of surgeon outcomes in the United Kingdom: potential caveats. *Int J Surg* 2014;12:369–70.
- 18 Chou E, Abboudi H, Shamim Khan M, *et al*. Should surgical outcomes be published? *J R Soc Med* 2015;108:127–35.
- 19 Williams MP, Modgil V, Drake MJ, *et al*. The effect of consultant outcome publication on surgeon behaviour: a systematic review and narrative synthesis. *Annals* 2018;100:428–35.
- 20 Radford PD, Derbyshire LF, Shalhoub J, *et al*. Publication of surgeon specific outcome data: a review of implementation, controversies and the potential impact on surgical training. *International Journal of Surgery* 2015;13:211–6.
- 21 Spaite DW, Bobrow BJ, Keim SM, *et al*. Association of statewide implementation of the prehospital traumatic brain injury treatment guidelines with patient survival following traumatic brain injury: the excellence in prehospital injury care (EPIC) study. *JAMA Surg* 2019;154:e191152.
- 22 O'Reilly GM, Curtis K, Kim Y, *et al*. Establishing determinants and quality indicators for getting home alive following moderate to severe traumatic brain injury: the Australian traumatic brain injury national data project. *Emerg Med Australas* 2021;33:1121–3.
- 23 Fountain DM, Davies SCL, Woodfield J, *et al*. Evaluation of nationwide referral pathways, investigation and treatment of suspected cauda equina syndrome in the United Kingdom. *Br J Neurosurg* 2019;33:624–34.
- 24 Williams M, Treasure P, Greenberg D, *et al*. Surgeon volume and 30 day mortality for brain tumours in England. *Br J Cancer* 2016;115:1379–82.
- 25 Lawrence T, Helmy A, Bouamra O, *et al*. Traumatic brain injury in England and Wales: prospective audit of epidemiology, complications and standardised mortality. *BMJ Open* 2016;6:e012197–9.
- 26 Fried HI, Nathan BR, Rowe AS, *et al*. The insertion and management of external ventricular drains: an evidence-based consensus statement: a statement for healthcare professionals from the neurocritical care society. *Neurocrit Care* 2016;24:61–81.
- 27 Clarke GM, Conti S, Wolters AT, *et al*. Evaluating the impact of healthcare interventions using routine data. *BMJ* 2019;7:l2239.
- 28 Lilford R, Mohammed MA, Spiegelhalter D, *et al*. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *The Lancet* 2004;363:1147–54.
- 29 Nepogodiev D, Omar OM, Glasbey JC, *et al*. Elective surgery cancellations due to the COVID-19 pandemic: global predictive modelling to inform surgical recovery plans. *Br J Surg* 2020;107:1440.
- 30 Schipmann S, Varghese J, Brix T, *et al*. Establishing risk-adjusted quality indicators in surgery using administrative data—an example from neurosurgery. *Acta Neurochir* 2019;161:1057–65.