Check for updates

Video-Based Performance Analysis in Pituitary Surgery—Part 1: Surgical Outcomes

Danyal Z. Khan[1,2], Chan Hee Koh[1,2], Adrito Das[2], Alexandra Valetopolou[1,2], John G. Hanrahan[1,2], Hugo Layard Horsfall[1,2], Stephanie E. Baldeweg[3,4], Sophia Bano[2], Anouk Borg[1], Neil L. Dorward[1], Olatomiwa Olukoya[1], Danail Stoyanov[2,5], Hani J. Marcus[1,2]

■ BACKGROUND: **Endoscopic pituitary adenoma surgery has a steep learning curve, with varying surgical techniques and outcomes across centers. In other surgeries, superior performance is linked with superior surgical outcomes. This study aimed to explore the prediction of patient-specific outcomes using surgical video analysis in pituitary surgery.**

■ METHODS: **Endoscopic pituitary adenoma surgery videos from a single center were annotated by experts for operative workflow (3 surgical phases and 15 surgical steps) and operative skill (using modified Objective Structured Assessment of Technical Skills [mOSATS]). Quantitative workflow metrics were calculated, including phase duration and step transitions. Poisson or logistic regression was used to assess the association of workflow metrics and mOSATS with common inpatient surgical outcomes.**

■ RESULTS: **100 videos from 100 patients were included. Nasal phase mean duration was 24 minutes and mean mOSATS was 21.2/30. Mean duration was 34 minutes and mean mOSATS was 20.9/30 for the sellar phase, and 11 minutes and 21.7/30, respectively, for the closure phase. The most common adverse outcomes were new anterior pituitary hormone deficiency (n = 26), dysnatremia (n = 24), and cerebrospinal fluid leak (n = 5). Higher mOSATS for all 3 phases and shorter operation duration were associated with decreased length of stay ($P = 0.003$ & $P < 0.001$). Superior closure phase mOSATS were associated with reduced postoperative cerebrospinal fluid leak ($P < 0.001$), and superior sellar phase mOSATS were associated with reduced postoperative visual deterioration ($P = 0.041$).**

■ CONCLUSIONS: **Superior surgical skill and shorter surgical time were associated with superior surgical outcomes, at a generic and phase-specific level. Such video-based analysis has promise for integration into data-driven training and service improvement initiatives.**

INTRODUCTION

Pituitary adenomas are among the most common intracranial tumors, with an estimated prevalence of up to 20%.[1,2] They are slow-growing tumors that may present incidentally through mass effect (e.g. visual decline) or hormone imbalance (e.g. Cushing's disease), therefore potentially causing significant morbidity, quality of life reduction, and death if left untreated.[1,3]

The gold standard treatment for most patients with symptomatic pituitary adenoma is transsphenoidal surgical excision, with the endoscopic transsphenoidal approach (eTSA) improving surgical access and visualization when compared with microscopic

Key words
- Outcomes
- Performance
- Pituitary adenoma
- Prediction
- Transsphenoidal

Abbreviations and Acronyms
CSF: Cerebrospinal fluid
eTSA: Endoscopic transsphenoidal approach
mOSATS: Modified Objective Structured Assessment of Technical Skills
OSATS: Objective Structured Assessment of Technical Skills
SIADH: Syndrome of inappropriate antidiuretic hormone

From the [1]Department of Neurosurgery, National Hospital for Neurology and Neurosurgery, [2]Wellcome / EPSRC Centre for Interventional and Surgical Sciences, University College London, [3]Department of Diabetes & Endocrinology, University College London Hospitals NHS Foundation Trust, [4]Division of Medicine, Department of Experimental and Translational Medicine, Centre for Obesity and Metabolism, University College London, and [5]Digital Surgery Ltd, Medtronic, London, UK

To whom correspondence should be addressed: Danyal Z. Khan, M.D. [E-mail: d.khan@ucl.ac.uk]

Danail Stoyanov and Hani J Marcus are denotes joint senior authorship with equal contribution.

Citation: World Neurosurg. (2024) 190:e787-e796.
https://doi.org/10.1016/j.wneu.2024.07.218

*Journal homepage: **www.journals.elsevier.com/world-neurosurgery***

*Available online: **www.sciencedirect.com***

approaches.[1,2] However, postoperative outcomes (e.g. dysnatremia) are challenging to predict after pituitary surgery, with patients often requiring days of monitoring postoperatively prior to safe discharge.[4] This inherently limits service improvement initiatives such as early discharge protocols or prophylactic treatment of common complications, for example,

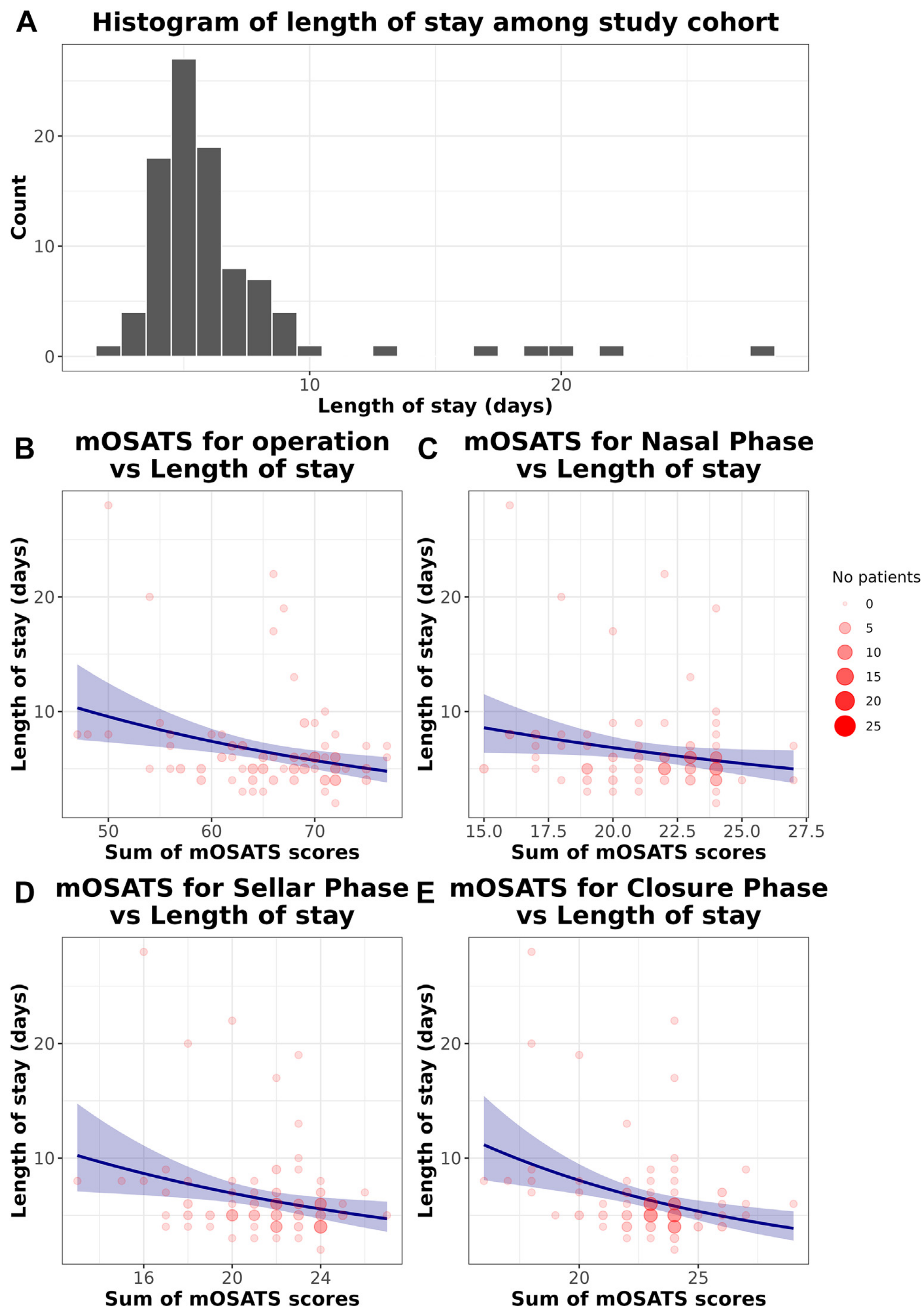

Figure 1. (**A**) Summarizes length of stay durations for the 100 patients. (**B–E**) mOSATS scores against length of stay. The logistic regression best fit is the blue line with a shaded 95% confidence interval. mOSATS, modified Objective Structured Assessment of Technical Skills.

prophylactic fluid restriction for those at high risk of syndrome of inappropriate antidiuretic hormone (SIADH).[5]

Patient and tumor-related factors likely affect such outcomes, for example, patient age, or tumor size and invasiveness.[6] While surgical performance has been shown to influence surgical outcomes in many procedures,[7] this has not been explored for eTSA. There is also significant variation in how the eTSA is performed, largely based on surgeon preference, and the impact of this on outcomes is unclear.[3,8-13] This variation contributes toward the challenge of auditing performance and predicting outcomes. To address this, we generated an international, consensus-driven analysis of the operative workflow (phases, steps, instruments, and errors) in contemporary pituitary surgery (**Figure 1**).[13] This workflow analysis is used to systematically segment operations into their subcomponents, which can then be assessed for their effectiveness and safety. Practically, this can be performed through the segmentation of operative videos, allowing a granular and structured analysis of surgical performance.[14-18] Similarly, operative skill can be assessed via video analysis using validated scales such as the Objective Structured Assessment of Technical Skills (OSATS), which we have modified and validated specifically for pituitary surgery (mOSATS).[16]

Linking this video-based workflow and skill analysis to the wider clinical context has the potential to allow data-driven postoperative outcome prediction based on surgical performance. In the future, this could be used to provide clinicians with postoperative decision support, identifying patients at low risk for complications who may be discharged early or those at high risk who would benefit from further monitoring or even prophylactic treatment. Furthermore, this video-based performance analysis could be automated using artificial intelligence and therefore integrated into surgical training programs to regularly evaluate and improve operative skill.

As a first step, we sought to explore the potential for surgical video performance analysis in predicting patient-specific outcomes after pituitary surgery.

METHODS

Study Overview

A retrospective analysis of prospectively collected operative video data and surgical outcome data was performed at a single tertiary neurosurgical center—the National Hospital for Neurology & Neurosurgery, Queen Square, London, United Kingdom. Ethical approval was granted for the project via the regional ethics committee of South West - Frenchay Research Ethics Committee, and informed written patient consent was obtained.

Data Collection and Curation

All endoscopic pituitary surgeries performed at our center are recorded and uploaded after patient consent to the Touch Surgery Enterprise platform (Medtronic, Dublin, Ireland), a combined software and hardware solution for securely recording, storing, and analyzing surgical videos. 100 consecutive videos of the eTSA for pituitary adenomas were selected from this library between August 16, 2018, and June 09, 2022. Incomplete videos or recent revision surgeries (within 6 months) were excluded. All videos were annotated for the surgical steps and phases present (**Table 1**), guided by a standardized annotation framework which was derived from a preceding international consensus study on pituitary surgery workflow.[13,14,19] Annotation was performed collaboratively by 2 neurosurgical residents with operative pituitary experience and then verified by an attending neurosurgeon. These step and phase annotations were used to calculate quantitative workflow metrics—individual step length, individual phase length, total number of step transitions, and total number of phase transitions. A transition was defined as when one step (or phase) changes to a different step (or phase).

All videos were then annotated for surgical performance using a bespoke mOSATS scale by 3 independent attending neurosurgeons based at external centers.[20] mOSATS scores were calculated for each of the 3 surgical phases (nasal, sellar, closure), with the 6 subdomains combined to give a total score for each phase (maximum score of 30 per phase) (**Supplementary Material 1**). For each video, the associated clinical data were extracted from a prospectively maintained database, with metrics derived from previous multicenter studies on pituitary surgery outcomes.[12,21] This included demographics (age, sex), tumor factors (size, subtype), and common inpatient outcomes—including length of stay, hyponatremia, SIADH, hypernatremia, diabetes insipidus (transient or persistent), cerebrospinal fluid (CSF) leak (biochemically confirmed and/or requiring operative intervention), postoperative visual deterioration, new suspected anterior pituitary hormone deficit requiring hydrocortisone supplementation on discharge (started if day 2 cortisol <350 nmol/L, except for patients with Cushing's), and new suspected posterior pituitary hormone deficit requiring desmopressin supplementation on discharge.

Table 1. Surgical Phases and Constituent Steps Labeled in the Surgical Videos if Present

Phases	Nasal	Sellar	Closure
Steps	1. Lateral displacement of middle and superior turbinates 2. Identification of sphenoid ostium and anterior sphenoidotomy 3. Lateral displacement of septum 4. Naso-sphenoid corridor creation 5. Identification of sella limits and neurovascular landmarks	6. Sellotomy 7. Durotomy 8. Tumor excision	9. Hemostasis 10. Spongostan placement 11. Fat graft placement 12. Rigid buttress placement 13. Fascia placement 14. Tissue glue application 15. Nasal packing

Data Analysis

Descriptive statistics were generated for the workflow, mOSATS and clinical data using Excel (v16.8, Microsoft). Workflow metrics and mOSATS were paired with particular outcomes on a phase level. Sellar phase mOSATS and workflow metrics were compared against outcomes that were felt to be relevant to that phase (hypernatremia, hyponatremia, new visual loss, new anterior or posterior pituitary hormonal deficit requiring hydrocortisone or desmopressin supplementation, respectively). Closure phase mOSATS and workflow metrics were compared against CSF leak rates. Nasal phase mOSATS and workflow metrics were compared against the length of stay only, as no outcomes in our dataset were felt to align logically with this phase.

Statistical analysis was conducted using R statistical programming,[22] with packages *dplyr*[23] and *tidyr*[24] for data preparation, and *ggplot2*[25] and *patchwork*[26] for data visualization. Continuous outcomes were analyzed with Poisson regression, and binary outcomes with logistic regression. Any "U"-shaped curves (or its inverse) were investigated by using a quadratic transformation of the data, with the peak/trough set to be the median of the data. The best fitting model for each outcome and metric combination was selected by using the Akaike Information Criterion.[27]

RESULTS

Overview

Data from 100 patients were included in this study. The median age of the sample was 53 years (interquartile range: 41–69) and the majority were male (n = 59). Most tumors were macroadenomas (n = 94), and the most common clinical phenotype was nonfunctioning adenoma (n = 71), followed by acromegaly (n = 16), prolactinoma (n = 7), and Cushing's disease (n = 6). The median length of stay was 5 days (IQR: 5–7 days; Figure 1), and the rest of the postoperative outcomes are displayed in Table 2. Of note, 7 of the 10 cases of diabetes insipidus were transient, and 1 displayed a triphasic response—developing SIADH. All 5 cases of CSF leaks required operative intervention (CSF diversion or skull base repair). Of the 2 patients with postoperative visual deterioration, one was due to a hematoma which needed surgical evacuation, while the other case was of presumed ischemic etiology.

Table 2. Summary of Postoperative Outcomes

Outcome	Number of Patients
Hyponatremia	13 (13%)
SIADH	11 (11%)
Hypernatremia	11 (11%)
Diabetes insipidus	10 (10%)
Cerebrospinal fluid leak	5 (5%)
Postoperative visual deterioration	4 (4%)
New anterior pituitary hormone deficit requiring hydrocortisone supplementation	26 (26%)
New posterior pituitary hormone deficit requiring desmopressin supplementation	3 (3%)

Seven of the 10 cases of diabetes insipidus were transient, and one displayed a triphasic response - developing SIADH.
SIADH, syndrome of inappropriate antidiuretic hormone.

Skill versus Outcomes

The mean overall mOSATS score across all 3 operative phases was 63.9 (standard deviation [SD]: 5.3), out of an available score of 90. For the nasal phase, the mean mOSATS score was 21.2 (SD: 1.9), out of an available score of 30. The mean score was lowest for the sellar phase at 20.9 (SD: 2.1) and highest for the closure phase at 21.7 (SD: 2.2). An example of high and low mOSATS scoring performances can be found in Supplementary Material 3.

When compared against postoperative outcomes, the mOSATS score for all 3 phases was significantly associated with decreased length of stay (Table 3, Figure 1). Additionally, a superior closure phase mOSATS score was significantly associated with a reduction in postoperative CSF leak incidence, and a superior sellar phase mOSATS was associated with a reduction in postoperative visual deterioration (Table 3, Figure 2).

Workflow Metrics versus Outcomes

The median phase duration was 24 minutes (IQR: 14–30 minutes) for the nasal phase, 34 minutes (IQR: 26–43 minutes) for the sellar phase, and 11 minutes (IQR: 7–19 minutes) for the closure phase. The longest surgical step was tumor excision (median 10 minutes, IQR 5–15 minutes), followed by sellotomy (median 5 minutes, IQR 3 – 9 minutes) and rigid buttress placement when present (median 4 minutes, IQR 3 – 7 minutes). The median number of step transitions per phase was 11 (IQR: 7–14) for nasal, 7 (IQR: 4–11) for sellar, and 6 (IQR: 3–8) for closure.

A shorter operation duration was significantly associated with a reduced length of stay, as was a shorter nasal phase duration (Table 3, Figures 3 and 4). No other significant relationships were found at the phase level (Table 3, Supplementary Material 2). When comparing surgical steps against relevant outcomes, no statistically or clinically significant linear relationship was detected (Supplementary Material 2).

DISCUSSION

Principal Findings

To our knowledge, this is the first study leveraging surgical video analysis of performance and workflow to predict outcomes in pituitary surgery.

We found that superior general operative performance (measured via mOSATS, across all 3 surgical phases) was linked with better overall outcomes, such as decreased postoperative length of stay. Similarly, when considering surgical workflow, progression through all surgical steps in a shorter time was significantly associated with a decreased length of postoperative stay. Furthermore, superior performance in specific surgical phases was related to clinically related outcomes. For example, superior closure phase skill was linked with reduced postoperative CSF leak rates—an outcome specific to this operative phase, and one of the most common complications of endoscopic pituitary surgery. Similarly, superior sellar phase skill (which includes

Table 3. Logistic Regression of Operative Skill and Workflow Metrics per Phase Against Outcomes

Phase	Domain	Outcome	*P* Value
Whole operation	Skill (mOSATS)	Length of stay§	**0.003**
	Duration	Length of stay†	**<0.001**
	No. of phase transitions	Length of stay‡	0.171
Nasal phase	Skill (mOSATS)	Length of stay§	**0.044**
	Duration	Length of stay‡	**0.037**
	No. of step transitions	Length of stay‡	0.063
Sellar phase	Skill (mOSATS)	Length of stay§	**0.013**
		Hypernatremia§	0.349
		Hyponatremia§	0.367
		Dysnatremia§	0.323
		Postoperative visual loss§	**0.041**
		New anterior pituitary hormonal deficit§	0.322
		New posterior pituitary hormonal deficit§	0.415
	Duration	Length of stay†	1.017
		Hypernatremia¶	0.638
		Hyponatremia¶	0.468
		Dysnatremia¶	0.511
		Postoperative visual loss¶	0.343
		New anterior pituitary hormonal deficit‖	0.398
		New posterior pituitary hormonal deficit§	0.397
	No. of step transitions	Length of stay*	0.191
		Hypernatremia¶	0.243
		Hyponatremia‖	0.533
		Dysnatremia¶	0.222
		Postoperative visual loss¶	0.282
		New anterior pituitary hormonal deficit‖	0.293
		New posterior pituitary hormonal deficit‖	0.996
Closure phase	Skill (mOSATS)	Length of stay§	**<0.001**
		Cerebrospinal fluid leak§	**0.025**
	Duration	Length of stay‖	0.056
		Cerebrospinal fluid leak‖	0.172
	No. of step transitions	Length of stay¶	0.419
		Cerebrospinal fluid leak¶	0.100

Dysnatremia is the sum of hypernatremia and hyponatremia. For each combination, best-fit models were selected. *P*-values surpassing the threshold of statistical significance are bolded mOSATS, modified Objective Structured Assessment of Technical Skills.
*Represents Poisson regression.
†Represents quadratic then Poisson regression.
‡Represents quadratic on log(x) then Poisson regression.
§Represents logistic regression.
‖Represents quadratic then logistic regression.
¶Represents quadratic on log(x) then logistic regression.

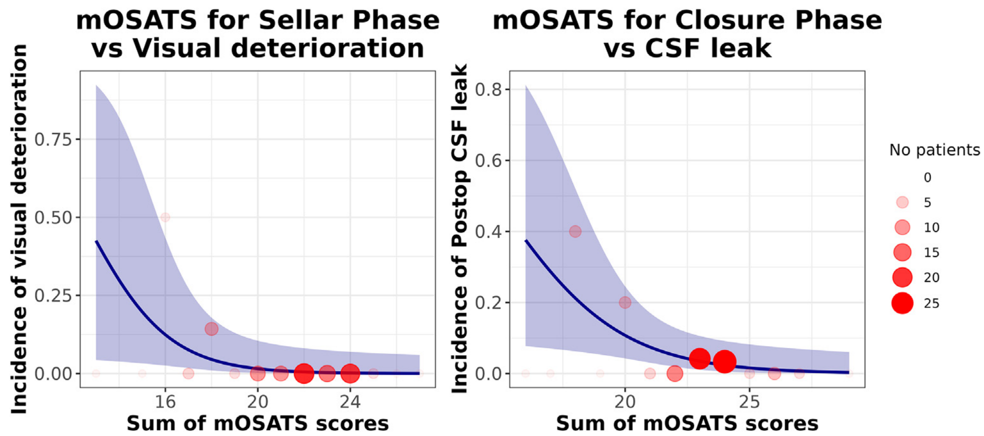

Figure 2. mOSATS scores for sellar phase and closure phase against postoperative visual deterioration and cerebrospinal fluid (CSF) leak, respectively. The logistic regression best fit is the blue line with a shaded 95% confidence interval. mOSATS, modified Objective Structured Assessment of Technical Skills.

tumor resection) was associated with reduced postoperative visual deterioration.

Despite the above relationships between operative skill and outcomes, workflow metrics at a more granular step level (for example, step length or step frequency) were not associated with specific outcomes. The utility of workflow metric analysis may be clearer in multicenter studies, with larger variations in operative performance.[28] Alternatively, analyzing workflow metrics that are even more granular, that is "actions" and "gestures," may uncover relationships between performance and outcomes which are not evident on a "phase" or "step" level, and may be more closely aligned with surgical skill.[28-30]

This video-based performance analysis sets the foundation for numerous avenues of clinical translation, with the ultimate goal of improving patient outcomes. Firstly, with regard to training, these quantitative metrics (e.g. workflow or skill assessment) can be presented to residents to help direct educational needs and supplement qualitative constructive feedback as part of coaching programs—akin to those seen in professional sports.[31] These metrics are linked with surgical outcomes, suggesting the potential clinical impact of such training interventions.[31,32] Similarly, this novel approach to surgical performance and outcome monitoring may have utility for clinical audit and raises the question of whether we should be recording every surgical procedure.[33] However, this requires significant

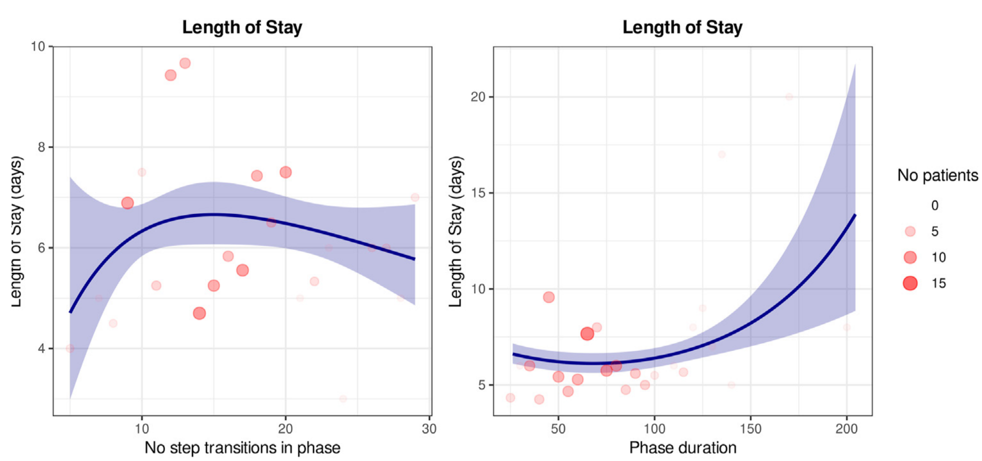

Figure 3. Whole operation workflow metrics against length of stay. The Poisson regression best fit is the blue line with a shaded 95% confidence interval.

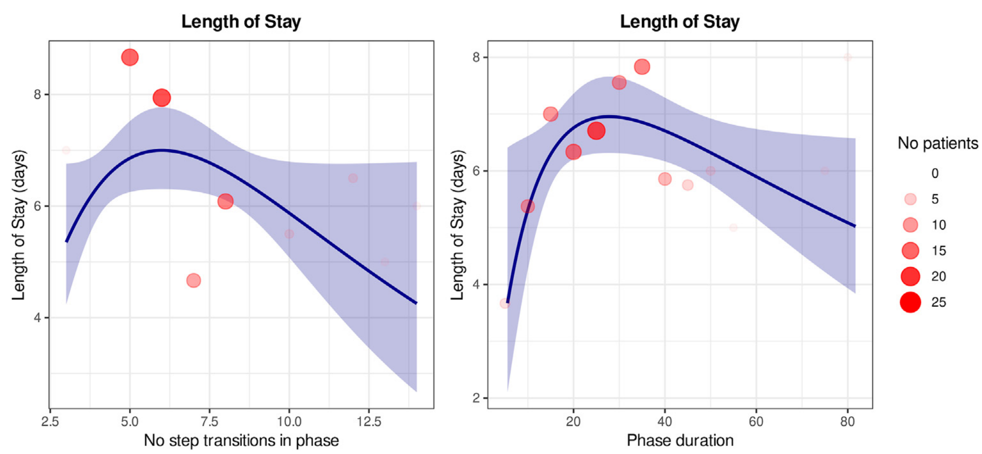

Figure 4. Nasal phase workflow metrics against length of stay. The Poisson regression best fit is the blue line with a shaded 95% confidence interval.

technological infrastructure at present for data storage and has uncertain medicolegal implications.

For these applications, this surgical performance analysis would ideally be more automated using technology such as computer vision and machine learning. Accordingly, our group has developed artificial neural networks for workflow recognition in pituitary surgery, which will improve the real-world feasibility and scalability of video analysis applications in the future.[33]

Finding in the Context of the Literature

Our findings dovetail with existing literature on the association between operative performance and generic surgical outcomes—heralded by Birkmeyer et al.'s study linking bariatric surgery video-based OSATS scores to early surgical outcomes and complications.[7] However, this relationship between surgical skill and outcomes did not persist for longer-term bariatric surgery outcomes, where other factors (e.g. patient-related) may have a stronger influence on outcomes (e.g. BMI, medical comorbidity resolution).[34] Otherwise, general surgery dominates the surgical video analysis literature—with a positive association between skills and outcomes found in laparoscopic colectomy surgery (using a bespoke skill assessment tool and OSATS) and laparoscopic total mesorectal excision (using a bespoke skill assessment tool), and laparoscopic gastrectomy for cancer (using OSATS for skill assessment), but not in less technically challenging procedures such as laparoscopic sleeve gastrectomy for weight loss (using OSATS for skill assessment).[32,35-37]

Considering pituitary surgery specifically, our previous systematic review of 193 articles exploring the modern practice of this closure phase of endoscopic pituitary surgery found absolute heterogeneity in operative techniques and resulting postoperative CSF leak rates.[11] Similarly, in a large multicenter observational study, we found a significant closure phase operative technique variation and CSF leak rates, with only a handful of factors influencing CSF leak rates (revision surgery, intraoperative CSF leak and tissue glue use).[21,38] The findings of this study suggest that operative skill is an important modifiable risk factor for CSF leak, and interventions at improving this are likely just as important as exploring the operative technique used and optimizing patient-related factors. Furthermore, our findings suggest higher operative skill during tumor resection is linked with a lower risk of postoperative visual deterioration—perhaps reflecting better optic apparatus decompression or more respect for visual structures during tumor removal. The literature suggests that visual deterioration after pituitary adenoma surgery is rare (1–2%), with the most common causes being compressive or ischemic.[39] Traditional risk factors include preoperative severe visual deficit, suprasellar tumor extension, and use of rigid repair materials but the influence of surgical skill certainly needs further study.[39,40]

More generally, many factors which influence outcomes after pituitary surgery go beyond the operation itself—for example, patient-related factors (e.g. age, frailty, prior surgery, comorbidities), tumor-related factors (e.g. size, invasiveness, histological subtype), and surgeon-related factors (e.g. surgical volume).[41-48] Therefore, although analysis of novel data types such as surgical videos may afford us the ability to better predict outcomes that are otherwise currently difficult to predict after pituitary surgery (e.g. SIADH), the future of robust outcome prediction models will be multimodal. This multimodal analysis will integrate clinical metadata, preoperative imaging, and operative videos in order to predict outcomes more accurately at a patient level.[33] Such data-driven outcome prediction may help form part of clinical decision support tools—aiding in the identification of patients at high risk of complication (who need further monitoring or prophylactic treatment) or those at low risk (who may be suitable for early discharge).[49,50] However, integrating these data will require advances in data analysis techniques, such as artificial intelligence, which allows automated analysis of novel datasets (e.g. images and videos) and can handle large amounts of variables with complex nonlinear relationships.[14,33,51] With these data science innovations and other complementary technological advances, the pituitary surgery pathway of the future may deliver more personalized and precise medicine.

Strengths and Limitations

The strengths of this work lie in the moderately sized dataset—representing the largest combined video and clinical outcome database published within endoscopic pituitary surgery thus far to our knowledge. Similarly, the data analysis was granular, and based on validated and evidence-based annotation frameworks. However, there are numerous limitations to this work. It is a single center, with workflow analysis limited to a phase and step level, and outcomes limited to early inpatient outcomes. The definition of some of these inpatient outcomes was pragmatic, in light of the heterogenous definitions seen in the literature. For example, the recording of anterior pituitary hormone dysfunction in this study was based on a new hydrocortisone supplementation requirement on discharge. Our center's threshold for starting this is relatively low early on postoperative (based on day 2 postoperative bloods), and many of these patients will not have a long-term hydrocortisone dependence on follow-up testing. Furthermore, reported dysnatraemia in this study encompassed any sodium abnormality, even a once off postoperative derangement (which can be confounded by many other factors, e.g. perioperative fluids), in the context of close sodium monitoring in our unit. Future work therefore will be multicenter, higher volume, with more granular video analysis, more clinical context (e.g. baseline characteristic data, radiological data), and standardized longitudinal outcomes.

CONCLUSION

In this single-center study, superior surgical skill and shorter surgical time were associated with superior generic surgical outcomes in endoscopic pituitary adenoma surgery. Better phase-specific skill (e.g. closure phase) was associated with better phase-specific outcomes (e.g. CSF leak). More granular workflow metrics were not associated with a difference in outcomes. Such video-based analysis has promise for integration into training programs to potentially improve skill and therefore outcomes further. Future work will involve larger multicenter multimodal datasets, with more granular video analysis, and automation using artificial intelligence.

CRediT AUTHORSHIP CONTRIBUTION STATEMENT

Danyal Z. Khan: Conceptualization, Formal analysis, Writing — original draft, Writing — review & editing. **Chan Hee Koh:** Formal analysis, Writing — review & editing. **Adrito Das:** Formal analysis, Visualization, Writing — review & editing. **Alexandra Valetopolou:** Data curation, Writing — review & editing. **John G. Hanrahan:** Data curation, Writing — review & editing. **Hugo Layard Horsfall:** Writing — review & editing. **Stephanie E. Baldeweg:** Writing — review & editing. **Sophia Bano:** Formal analysis, Writing — review & editing. **Anouk Borg:** Conceptualization, Data curation, Writing — review & editing. **Neil L. Dorward:** Conceptualization, Data curation, Writing — review & editing. **Olatomiwa Olukoya:** Data curation, Writing — review & editing. **Danail Stoyanov:** Supervision, Writing — review & editing. **Hani J. Marcus:** Supervision, Writing — review & editing.

ACKNOWLEDGMENTS

With thanks to Digital Surgery Ltd., a Medtronic company, for access to Touch SurgeryTM Enterprise for both video recording and analysis throughout this study.

REFERENCES

1. Ezzat S, Asa SL, Couldwell WT, et al. The prevalence of pituitary adenomas. *Cancer*. 2004;101:613-619.

2. Asa SL. Practical pituitary pathology: what does the pathologist need to know? *Arch Pathol Lab Med*. 2008;132:1231-1240.

3. Buchfelder M, Schlaffer S. Pituitary surgery for Cushing's disease. *Neuroendocrinology*. 2010; 92(Suppl 1):102-106.

4. Hussein Z, Tzoulis P, Marcus HJ, et al. The management and outcome of hyponatraemia following transsphenoidal surgery: a retrospective observational study. *Acta Neurochir (Wien)*. 2022; 164:1135-1144.

5. Yu S, Taghvaei M, Collopy S, et al. Evaluation of early postoperative day 1 discharge after endoscopic endonasal pituitary adenoma resection. *J Neurosurg*. 2022;136:1337-1346.

6. Lobatto DJ, de Vries F, Zamanipoor Najafabadi AH, et al. Preoperative risk factors for postoperative complications in endoscopic pituitary surgery: a systematic review. *Pituitary*. 2018; 21:84-97.

7. Birkmeyer JD, Finks JF, O'Reilly A, et al. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013;369:1434-1442.

8. Lucas JW, Zada G. Endoscopic surgery for pituitary tumors. *Neurosurg Clin*. 2012;23:555-569.

9. Shah NJ, Navnit M, Deopujari CE, Mukerji SS. Endoscopic pituitary surgery - a. beginner's guide. *Indian J Otolaryngol Head Neck Surg*. 2004;56:71-78.

10. Cappabianca P, Cavallo LM, de Divitiis O, Solari D, Esposito F, Colao A. Endoscopic pituitary surgery. *Pituitary*. 2008;11:385-390.

11. Khan DZ, Ali AMS, Koh CH, et al. Skull base repair following endonasal pituitary and skull base tumour resection: a systematic review. *Pituitary*. 2021;24:698-713.

12. CRANIAL Consortium. CSF rhinorrhoea after endonasal intervention to the skull base (CRANIAL) - Part 1: multicenter pilot study. *World Neurosurg*. 2021;149:e1077-e1089.

13. Marcus HJ, Khan DZ, Borg A, et al. Pituitary society expert Delphi consensus: operative workflow in endoscopic transsphenoidal pituitary adenoma resection. *Pituitary*. 2021;24:839-853.

14. Khan DZ, Luengo I, Barbarisi S, et al. Automated operative workflow analysis of endoscopic pituitary surgery using machine learning: development and preclinical evaluation (IDEAL stage 0). *J Neurosurg*. 2022;137:51-58.

15. Zisimopoulos O, Flouty E, Luengo I, et al. DeepPhase: surgical phase recognition in CATARACTS videos. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, eds. *Medical Image Computing and Computer Assisted Intervention — MICCAI 2018*. Cham: Springer International Publishing; 2018:265-272.

16. Garrow CR, Kowalewski KF, Li L, et al. Machine learning for surgical phase recognition: a systematic review. *Ann Surg*. 2021;273:684-693.

17. Hashimoto DA, Rosman G, Rus D, Meireles OR. Artificial intelligence in surgery: promises and perils. *Ann Surg*. 2018;268:70-76.

18. Hashimoto DA, Rosman G, Witkowski ER, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg*. 2019;270: 414-421.

19. Das A, Bano S, Vasconcelos F, Khan DZ, Marcus HJ, Stoyanov D. Reducing Prediction volatility in the surgical workflow recognition of endoscopic pituitary surgery. *Int J Comput Assist Radiol Surg*. 2022;17:1445-1452.

20. Newall N, Khan DZ, Hanrahan JG, et al. High fidelity simulation of the endoscopic transsphenoidal approach: validation of the UpSurgeOn TNS Box. *Front Surg*. 2022;9:1049685.

21. CRANIAL Consortium. CSF rhinorrhoea after endonasal intervention to the skull base (CRANIAL): a multicentre prospective observational study. *Front Oncol*. 2022;12:1049627.

22. R: The R Project for Statistical Computing [Internet]. R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2018.

23. Wickham H, François R, Henry L, Davis V. Dplyr: a grammar of data manipulation Inside Internet. Available at: https://CRAN.R-project.org/package=dplyr; 2023. Accessed February 1, 2024.

24. Wickham H, Davis V, Girlich M. Tidyr: tidy messy data. R package version 1.3.0. [Internet]. Available at: https://CRAN.R-project.org/package=tidyr; 2023. Accessed February 1, 2024.

25. Wickham H. *Elegant Graphics for Data Analysis*. New York: Springer-Verlag; 2016.

26. Pedersen TL. Patchwork: the composer of plots [internet]. Available at: https://CRAN.R-project.org/package=patchwork; 2022. Accessed February 1, 2024.

27. Vrieze SI. Model selection and psychological theory: a discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychol Methods*. 2012;17:228-243.

28. Ma R, Ramaswamy A, Xu J, et al. Surgical gestures as a method to quantify surgical performance and predict patient outcomes. *NPJ Digit Med*. 2022;5: 187.

29. Meireles OR, Rosman G, Altieri MS, et al. SAGES consensus recommendations on an annotation framework for surgical video. *Surg Endosc*. 2021;35: 4918-4929.

30. Chen AB, Liang S, Nguyen JH, Liu Y, Hung AJ. Machine learning analyses of automated performance metrics during granular sub-stitch phases predict surgeon experience. *Surgery*. 2021;169: 1245-1249.

31. Daniel R, McKechnie T, Kruse CC, et al. Video-based coaching for surgical residents: a systematic review and meta-analysis. *Surg Endosc*. 2023;37: 1429-1439.

32. Stulberg JJ, Huang R, Kreutzer L, et al. Association between surgeon technical skills and patient outcomes. *JAMA Surgery*. 2020;155:960-968.

33. Khan DZ, Hanrahan JG, Baldeweg SE, Dorward NL, Stoyanov D, Marcus HJ. Current and future advances in surgical therapy for pituitary adenoma. *Endocr Rev*. 2023;44:947-959.

34. Scally CP, Varban OA, Carlin AM, Birkmeyer JD, Dimick JB. For the Michigan bariatric surgery collaborative. Video ratings of surgical skill and late outcomes of bariatric surgery. *JAMA Surgery*. 2016;151:e160428.

35. Brajcich BC, Stulberg JJ, Palis BE, et al. Association between surgical technical skill and long-term survival for colon cancer. *JAMA Oncol*. 2021;7: 127-129.

36. Varban OA, Thumma JR, Finks JF, Carlin AM, Ghaferi AA, Dimick JB. Evaluating the effect of surgical skill on outcomes for laparoscopic sleeve gastrectomy: a video-based study. *Ann Surg*. 2021; 273:766.

37. Fecso AB, Bhatti JA, Stotland PK, Quereshy FA, Grantcharov TP. Technical performance as a predictor of clinical outcomes in laparoscopic gastric cancer surgery. *Ann Surg*. 2019;270:115.

38. Machine learning driven prediction of cerebrospinal fluid rhinorrhoea following endonasal skull base surgery: a multicentre prospective observational study. *Front Oncol*. 2023;13:1046519.

39. Carnevale JA, Babu CS, Goldberg JL, Fong R, Schwartz TH. Visual deterioration after endonasal endoscopic skull base surgery: causes, treatments, and outcomes. *J Neurosurg*. 2021;136:1103-1113.

40. Rutland JW, Dullea JT, Oermann EK, et al. Postoperative vision loss: analysis of 587 patients undergoing endoscopic surgery for pituitary macroadenoma. *Br J Neurosurg*. 2022;36:494-500.

41. Mooney MA, Sarris CE, Zhou JJ, et al. Proposal and validation of a simple grading scale (TRANSSPHER grade) for predicting gross total resection of nonfunctioning pituitary macroadenomas after transsphenoidal surgery. *Operative Neurosurgery*. 2019;17:460-469.

42. Zanier O, Zoli M, Staartjes VE, et al. Machine learning-based clinical outcome prediction in surgery for acromegaly. *Endocrine*. 2022;75:508-515.

43. Hollon TC, Parikh A, Pandian B, et al. A machine learning approach to predict early outcomes after pituitary adenoma surgery. *Neurosurg Focus*. 2018; 45:E8.

44. Stroud A, Dhaliwal P, Alvarado R, et al. Outcomes of pituitary surgery for Cushing's disease: a systematic review and meta-analysis. *Pituitary*. 2020; 23:595-609.

45. Thommen R, Kazim SF, Cole KL, et al. Worse pituitary adenoma surgical outcomes predicted by increasing frailty, not age. *World Neurosurg*. 2022; 161:e347-e354.

46. Fang Y, Wang H, Feng M, et al. Machine-learning prediction of postoperative pituitary hormonal outcomes in nonfunctioning pituitary adenomas: a multicenter study. *Front Endocrinol (Lausanne)*. 2021;12:748725.

47. Buchy M, Lapras V, Rabilloud M, et al. Predicting early post-operative remission in pituitary adenomas: evaluation of the modified knosp classification. *Pituitary*. 2019;22:467-475.

48. Chalif EJ, Couldwell WT, Aghi MK. Effect of facility volume on giant pituitary adenoma neurosurgical outcomes. *J Neurosurg*. 2022;137:658-667.

49. Perez-Vega C, Tripathi S, Domingo RA, et al. Fluid restriction after transsphenoidal surgery for the prevention of delayed hyponatremia: a systematic review and meta-analysis. *Endocr Pract*. 2021;27:966-972.

50. Hughes MA, Culpin E, Darley R, et al. Enhanced recovery and accelerated discharge after endoscopic transsphenoidal pituitary surgery: safety, patient feedback, and cost implications. *Acta Neurochir (Wien)*. 2020;162:1281-1286.

51. Das A, Khan DZ, Hanrahan JG, Marcus HJ, Stoyanov D. Automatic generation of operation notes in endoscopic pituitary surgery videos using workflow recognition. *Intell Based Med*. 2023;8: 100107.

Conflict of interest statement: This work was supported in whole, or in part, by the Wellcome/EPSRC Centre for Interventional and Surgical Sciences (WEISS) [203145/Z/16/Z], the Engineering and Physical Sciences Research Council (EPSRC) [EP/W00805X/1, EP/Y01958X/1], the Department of Science, Innovation and Technology (DSIT), and the Royal Academy of Engineering under the Chair in Emerging Technologies programme. D. Z. Khan is supported by a National Institute for Health and Care Research (NIHR) Academic Clinical Fellowship and a Cancer Research UK Predoctoral Fellowship. A. Das is supported by the (EPSRC) [EP/S021612/1]. H. J. Marcus is supported by WEISS [NS/A000050/1] and by the NIHR Biomedical Research Centre at University College London (UCL). For the purpose of Open Access, the authors have applied a CC BY public copyright license to any author accepted manuscript version arising from this submission. D. Stoyanov is an employee of Digital Surgery, Medtronic, which is developing products related to the research described in this paper. H. Marcus is employed by Panda Surgical and holds shares in the company.

Presented via oral podium presentation at the Pituitary Clinicopathological Conference 2024 (London).

Received 14 July 2024; accepted 31 July 2024

Citation: World Neurosurg. (2024) 190:e787-e796.
https://doi.org/10.1016/j.wneu.2024.07.218

Journal homepage: ***www.journals.elsevier.com/world-neurosurgery***

Available online: ***www.sciencedirect.com***

Supplementary Table 1. mOSATS scoring scale

Phase and Domain	Scale
Nasal phase	
Respect for tissue	1 (poor) — 5 (perfect)
Time and motion	1 (poor) — 5 (perfect)
Instrument handling	1 (poor) — 5 (perfect)
Flow of operation	1 (poor) — 5 (perfect)
Knowledge of instruments	1 (poor) — 5 (perfect)
Knowledge of procedure	1 (poor) — 5 (perfect)
Sellar phase	
Respect for tissue	1 (poor) — 5 (perfect)
Time and motion	1 (poor) — 5 (perfect)
Instrument handling	1 (poor) — 5 (perfect)
Flow of operation	1 (poor) — 5 (perfect)
Knowledge of instruments	1 (poor) — 5 (perfect)
Knowledge of procedure	1 (poor) — 5 (perfect)
Closure phase	
Respect for tissue	1 (poor) — 5 (perfect)
Time and motion	1 (poor) — 5 (perfect)
Instrument handling	1 (poor) — 5 (perfect)
Flow of operation	1 (poor) — 5 (perfect)
Knowledge of instruments	1 (poor) — 5 (perfect)
Knowledge of procedure	1 (poor) — 5 (perfect)

mOSATS, modified Objective Structured Assessment of Technical Skills.