Langenbeck's Archives of Surgery (2019) 404:421–430
https://doi.org/10.1007/s00423-019-01798-7

ORIGINAL ARTICLE

Effect of surgeons' annual operative volume on the risk of permanent Hypoparathyroidism, recurrent laryngeal nerve palsy and Haematoma following thyroidectomy: analysis of United Kingdom registry of endocrine and thyroid surgery (UKRETS)

S. Aspinall[1] · D. Oweis[2] · D. Chadwick[3]

Received: 23 June 2018 / Accepted: 6 May 2019 / Published online: 28 June 2019

Abstract

Purpose Categorize data to investigate the surgeon volume outcome relationship in thyroidectomies. Determine the evidence base for recommending a minimum number of thyroidectomies performed per year to maintain surgical competency.

Methods Data on thyroid operations in the United Kingdom Registry of Endocrine and Thyroid Surgery (UKRETS) from 01/09/2010 to 31/08/2016 was analysed. The primary outcome measure was permanent hypoparathyroidism (PH). Recurrent laryngeal nerve palsy (RLN) and post-operative haematoma were also examined. Exclusion criteria included patient age > 85 or < 18 years, and surgeons contributing <10 operations. Data analysis was performed using general additive models and mixed effect logistic regression for PH and binary logistic regression for others.

Results For PH 10313 bilateral thyroid operations were analysed. The Annual rate (AR, $p = 0.012$) and nodal dissection ($P < 10^{-7}$) were significant factors. 25,038 thyroidectomies were analysed to investigate the effect of surgeon Volume on RLN palsy and haematoma. Age, retrosternal goitre, routine laryngoscopy, re-operation, nodal Dissection, bilateral thyroidectomy, RLN monitoring and surgeon volume were significantly associated with RLN palsy. Post-operative haematoma showed no significant correlation to surgeon volume. Categorisation of AR showed that PH and RLN palsy rates declined in surgeons performing >50 cases/year to a minimum of 3% and 2.6% respectively in highest volume AR group (>100 cases/year).

Conclusion Surgeon annual operative volume is a factor in determining outcome from thyroid surgery. Results are limited by a high proportion of missing data, which could potentially bias the outcome, but tentatively suggests the minimum recommended number of thyroid operations / year should be 50 cases.

Keywords Endocrine surgery · Surgeon volume · Thyroidectomy · Volume-outcome · Chronic hypocalcaemia, permanent hypoparathyroidism, operative volume

Introduction

There is a wealth of published evidence linking surgeon volume and outcome, with high volume surgeons achieving lower mortality, shorter hospital stay and fewer complications across a number of surgical specialties. Explanations for this effect include "practice makes perfect" and "selective referral" theories [1]. This has influenced clinical practice, with more specialization and high-risk procedures being increasingly undertaken by high-volume surgeons [2].

Improved outcomes in high volume surgeons are also seen in endocrine surgery, with fewer complications, decreased length of stay (LOS) and reduced hospital charges [3]. Likewise in thyroid surgery a number of studies published over the past 20 years have shown improvements in the thyroid surgery specific complications of hypocalcaemia, recurrent laryngeal nerve palsy (RLN) and post-operative haematoma, as well as LOS and general post-operative complications with increasing surgeon case-load [4–12]. Table 1.

✉ S. Aspinall
Sebastian.Aspinall@nhs.net

1 Consultant Endocrine Surgeon, NHS Grampian, Aberdeen Royal Infirmary, Aberdeen, UK

2 Specialist Registrar in General Surgery, Northumbria Healthcare NHS trust, North Tyneside General Hospital, North Shields, UK

3 Consultant Endocrine Surgeon, Nottingham University Hospitals NHS Trust, City Campus, Nottingham, UK

Table 1 Summary of Results of Studies investigating relationship between Surgeon Volume and Outcome in Thyroidectomy [4–12]

Author	Year of Publication	Country	Number	Source of Database	Study Period	Hypocalcaemia	RLN palsy	Haematoma	Length of Stay	General	Low volume	Intermediate volume	High volume	Very high volume
Sosa	1998	USA	5860	Maryland Hospital Discharge	1991–1996		yes #			yes	1–9	10–29	30–100	>100
Dralle	2004	Germany	16448	Prospective	1998–2001		yes				<45*		>45*	
Gourin	2010	USA	21270	Maryland Hospital Discharge	1990–2009	yes	yes		yes		<3	4–24	>24	
Kandil	2013	USA	21625	National Inpatient Sample	2000–2009	yes	yes	yes		yes	<10	10–99	>100	
Loyo	2013	USA	871644	National Inpatient Sample	1993–2008	yes	yes			yes	<3	4–9	9–23	>23
Adkisson	2014	USA	1249	Pittsburgh Medical Centre	2011	yes	yes	yes			<30		>30	
Hauch	2014	USA	62722	National Inpatient Sample	2003–2009	yes		yes	yes	yes	<10	10–99	>99	
Nouraei	2017	UK	72594	Hospital Episode Statistics	2004–2012		yes		yes		<10	10–50	>50	
Adam	2017	USA	16954	National Inpatient Sample	1998–2009	yes	yes	yes	yes	yes	<25		>25	

yes denotes study found improved outcome for variable with high volume surgeon

*Recurrent laryngeal nerves at risk

The source of data in all but two of these thyroid volume-outcome studies is hospital administrative datasets [5, 9]. These datasets are derived from inpatient or discharge information and are prone to subjectivity, variability and error which limit their accuracy, and ultimately affect the quality of the results [13]. Data in these databases is restricted to inpatient activity, and complications are recorded retrospectively, using international classifications of disease (ICD) codes, assigned by non-clinical coders, often without precise definition, but on the basis of whether these complications have been recorded in the case notes, which leads to underestimation of complication rates [13].

For example the rate of post-operative transient hypocalcaemia is 5.2% in a study from the United Kingdom, using Hospital Episode Statistics (HES) [11]. Whereas the United Kingdom Registry of Endocrine and Thyroid Surgeons (UKRETS), which is compiled by clinicians reports the incidence of temporary post-operative hypocalcaemia as 27% [14].

UKRETS is a national database which collects information on elective and emergency thyroid, parathyroid, adrenal and endocrine pancreas operations. It was started by members of the British Association of Endocrine and Thyroid Surgeons (BAETS) in 2004. The database contains preoperative, intra-operative and post-operative clinical data, as well as pathology and complications. A National Audit Report is published every 3 years from the data in UKRETS, which is currently in its 5th edition [14]. The data entry fields for UKRETS are outlined in the appendix to its report. UKRETS is mandatory for surgeons in England but it also collects data for surgery in England, Wales, Northern Ireland (NI) and Scotland. The data is collected by BAETS members and is not validated.

Clinicians enter data into the database contemporaneously or retrospectively for each surgical operation via an online portal which is managed by Dendrite Clinical Systems Limited on behalf of the BAETS. The database currently holds data on more than 100,000 endocrine operations. 19,395 thyroid and parathyroid operations were recorded in HES [15] in England over 12 months from April 2016. 10,631 operations were recorded into UKRETS over the same time period, which represents 55% (10,631/19395). But as UKRETS includes data not only from England but also from NI, Scotland and Wales, then the total proportion of thyroid and parathyroid operations in UK added to UKRETS will be less than 55%. However, data from England does form the vast majority of UKRETS data.

Although the relationship between surgeon volume and outcome is established, there is no accepted threshold to define a 'high volume' surgeon, nor method to determine its value [1]. Consequently the volume outcome thyroid studies each have different definitions of high volume surgery. The minimum number of thyroidectomies performed per year to maintain competency in the UK is currently 20 [16], though there isn't a clear evidence base to support this particular number of cases.

In this study UKRETS has been analysed to determine the evidence base for recommending a minimum of 20 thyroidectomies per year to maintain competency and a logical method for categorizing data into groups to analyse the surgeon volume outcome. This has been investigated primarily through the effect of surgeon's annual thyroidectomy case-load on permanent hypoparathyroidism (PH). RLN palsy and post-operative bleeding have been also investigated.

Methods

The data in the UKRETS over 6 years from 1st September 2010 to 31st August 2016 was analysed. In 2009 hypocalcaemia definitions were amended and strict criteria used to improve data entry, hence the time period starting in 2010 in this study.

From the available data in UKRETS on the three common complications of thyroidectomy i.e. hypocalcaemia, RLN palsy and post-operative haematoma; late hypocalcaemia or permanent hypoparathyroidism (PH) was considered to be the best quality indicator of thyroidectomy for reasons outlined in the discussion.

This study includes two separate analyses of UKRETS undertaken on extracts of the database from 2017 and 2018. Analysis of PH following thyroidectomy was examined in more depth than RLN palsy or post-operative haematoma.

Analysis of permanent hypoparathyroidism

PH is defined in UKRETS as the need to take calcium and/or vitamin D supplements to maintain normocalcaemia at 6 months following thyroidectomy. Cases with missing data for this outcome were excluded from the analysis. Predictive factors for PH included surgeons' annual operation rate (AR), patients' age, gender, identity of surgeon (to examine effect of clustering), and nodal dissection. These factors had been identified to be associated with PH from multivariate logistic regression analysis published in the 2012 National Audit report [17]. Table 2.

Surgeon annual operation rate was calculated by dividing the number of thyroid operations performed by the length of time the surgeon contributed cases. The effect of Graves' disease on the probability of PH was not analysed, as it wasn't significantly associated with PH in the 2012 analysis [17]. Both thyroid cancer and nodal dissection although shown to influence the probability of PH, are also independently correlated, and as nodal dissections are undertaken exclusively for thyroid cancer, it was illogical to include both in the analysis, and nodal dissection alone was analysed.

Table 2 Multivariate logistic regression analysis of predictive factors for late hypocalcaemia/PH following bilateral thyroid surgery according to the 4th National audit report UKRETS [17]

	OR (95% CI)	*P* value
Age	0.995 (0.990–1.001)	0.081
Gender	0.81 (0.651–1.009)	0.060
Graves' Disease	1.031 (0.842–1.262)	0.770
Thyroid cancer	1.462 (1.095–1.952)	0.010
Node Dissection	1.758 (1.294–2.390)	0.0003

Unilateral cases were excluded from PH analysis as this is unusual following first-time unilateral thyroid surgery [14, 17]. Patients who died within 6 months of surgery or who were lost to follow-up were also excluded. Patients aged >85 and < 18 years - who compromised the minority - were excluded, as there was some concern regarding the quality of data in UKRETS at the extremes of age, these represent a different subset e.g. paediatric group. A single statistical model accommodating such wide age range was considered unwise. Surgeons who had reported <10 cases in total to UKRETS over the study period were excluded, as these were likely to have started contributing to the database towards the end, or stopped entering data near the start of the study period, and so would have an inappropriate AR.

RLN palsy & post-operative haematoma

Predictive factors examined in UKRETS for analysis of RLN palsy and re-operation for post-operative bleeding included goitre type, re-operation, operative extent (unilateral versus bilateral surgery, node dissection), use of nerve monitor, pathology, patient age, gender, thyroid status at presentation, use of routine laryngoscopy and energy devices in addition to monopolar / bipolar diathermy. Cases with missing data in any of these data entry fields were excluded for analysis of RLN palsy / haematoma.

The same exclusion criteria with regard to patient age and surgeons recording <10 cases in total were applied. RLN palsy was defined as the occurrence of an abnormal vocal cord at post-operative laryngoscopy. Re-operation for bleeding is a binary value data entry field in UKRETS. Routine laryngoscopy was defined as any patient undergoing both pre and post-laryngoscopy.

Statistical modelling

Logistic regression was used to determine how the experience of the surgeon, represented by the surgeon's AR, affected the chance of the patient suffering from PH, re-operation for haematoma or RLN palsy. Analysis of PH was undertaken using R software on a 2017 extract from UKRETS. The analysis of RLN palsy and re-operation for haematoma was done using SPSS software from a 2018 extract.

To assess the assumption that PH declined with surgeon AR, this variable was categorized, so that a different risk was ascribed to each category. Generalized Additive Models (GAMs), which models the effect of a variable with a smooth rather than straight line, was chosen to as an interim analysis to examine this, and the GAMs curve generated was then used to inform the choice of categorisation of AR.

In the in depth analysis of PH, the outcome of each operation performed wasn't assumed to be independent, and a correction for dependence of outcome among cases from the same surgeon was therefore carried out. Mixed effect logistic regression (GLMMs), which models the effect between surgeons explicitly, was chosen for this analysis.

Results

Analysis of permanent hypoparathyroidism

38,422 thyroid operations were recorded in 2017 extract from UKRETS by 253 surgeons. 25,548 patients underwent unilateral thyroidectomy; 49 were lost to follow-up or died within 6 months; 304 patients had a recorded age < 18 or > 85 years; and 36 cases were performed by surgeons contributing <10 cases overall. After excluding these cases there were 12,485 bilateral thyroidectomies, performed by 218 surgeons. Missing data on PH was found in 2172 (17.4%) of these 12,485 cases, leaving 10,313 to analyse (Fig. 1).

Surgeons performed between 4.7 and 154.8 thyroidectomies per year. Patient gender did not affect the probability of PH ($p = 0.14$). Patient age ($p = 0.022$), surgeon volume ($p = 10^{-3}$) and nodal dissection ($p = 10^{-12}$) significantly affected PH in the GAMs analysis, without accounting for dependence of cases. Figures 2 shows the results of GAMs to model the effects of surgeon AR against the probability of PH (represented in on the vertical axis as a logistic scale).

Surgeons AR was classed into groups: < 25 cases/year, 25–50 cases/year, 50–75 cases/year, 75–100 cases/year and > 100 cases/year, as these categories reflect the shape of the curve in Fig. 2 and are also readily interpretable as representing a frequency of performing a thyroidectomy incrementally from less than fortnightly to more than twice a week. The numbers of BAETS surgeons falling into these annual rate categories is shown in Fig. 3.

Analysis of patients' age and surgeon AR with mixed effect logistic regression (to take into account clustering of cases), following categorisation of these variables, and with central nodal dissection as a binary variable, showed that patient age lost significance ($p = 0.12$) for predicting the probability of PH, but the effect of surgeon AR on the probability of PH retained significance ($p = 0.012$) and is shown in Table 3 with median, lower and upper quartile ranges. The lowest (<25 cases/year) volume surgeons group have a lower incidence of PH (6%) than the intermediate (25–50 cases/year) volume group

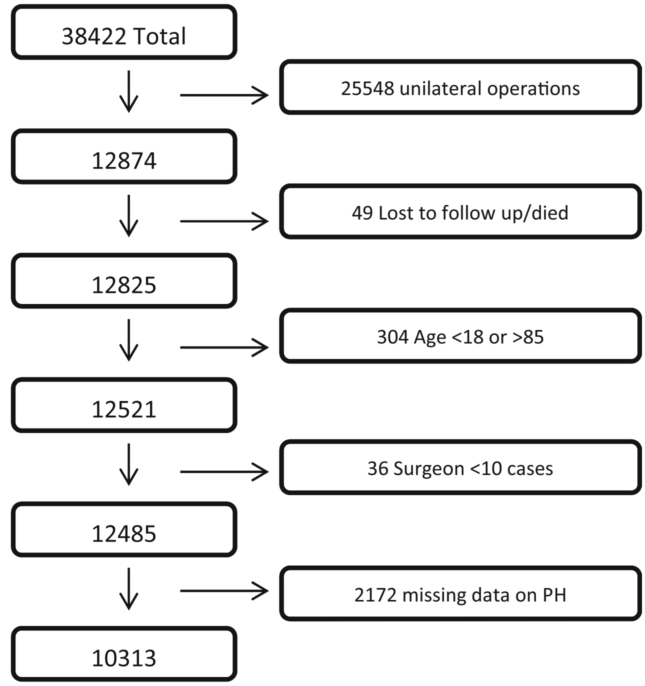

Fig. 1 Number of thyroidectomy operations analysed for probability of PH following application of exclusion criteria

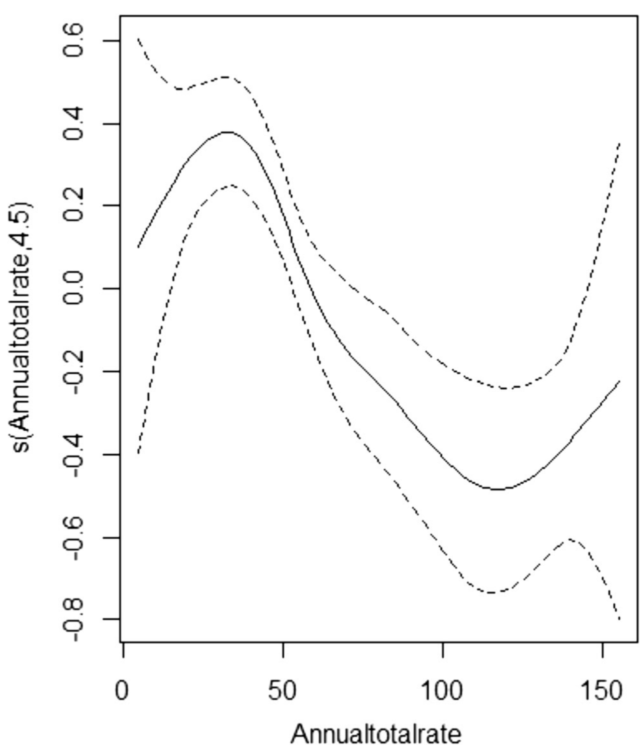

Fig. 2 Generalized Additive Model fits for probability of permanent hypoparathyroidism (logistic scale vertical axis) against surgeon annual volume – smoothed contributions (solid line) and 95% confidence intervals (dashed line)

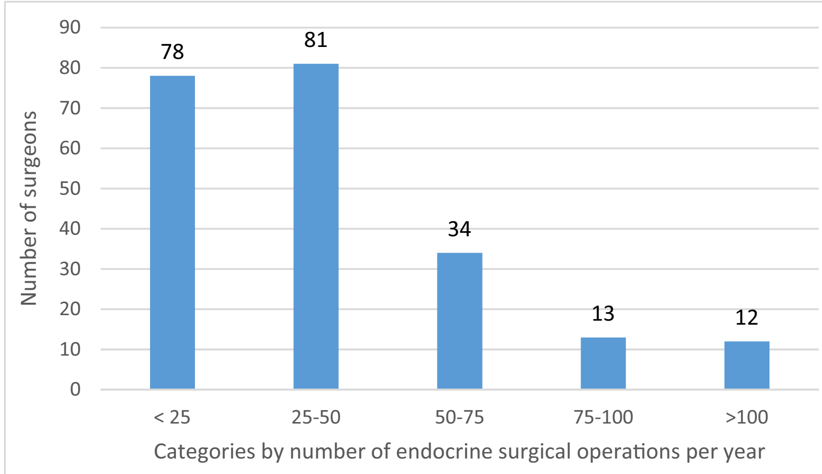

Fig. 3 Distribution of British Association of Endocrine and Thyroid Surgeons across categories of number of endocrine surgical operations performed per year

(6.6%). The risk of PH declines thereafter reaching a minimum of 3% in the highest (>100 cases/year) volume group.

The effect of nodal dissection on probability of PH is also maintained ($p < 10^{-8}$). Findings are similar to the no nodal dissection group (though with a higher incidence of PH, as expected), with an increase in the probability of PH from the lowest annual rate group (15.1%) to the intermediate annual rate group (16.6%), and a decline thereafter to a risk of PH of 8% in the highest surgeon volume category.

Data on PH was missing overall in 8139 (21.2%) of the 38,422 operations recorded in UKRETS. Analysis was done to determine whether the probability of this data being missing depended in a systematic way on the surgeon AR and if so could potentially bias the outcome of this study. A marginal effect of AR on the probability of absent PH data was found ($p = 0.039$), though the rate of absent data across groups was quite uniform, except in the 75–100 cases/year group (Table 4).

Analysis of RLN palsy and post-operative haematoma

The 2018 UKRETS identified a total of 39,659 thyroidectomies, indicating that an additional 1237 operations had been added retrospectively to the database since the 2017 extract. 757 cases were undertaken in patients aged <18 and > 85 years; 139 cases were performed by surgeons who contributed <10 cases overall; and 13,725 cases had missing data in the predictive factors examined. After applying these exclusion criteria 25,038 operations were included in the analysis (Fig. 4).

RLN palsy occurred in 963 / 25,038 (3.8%) cases overall in the time period examined. 25,038 thyroidectomies included 15,438 unilateral procedures and the remaining bilateral. There were 963 RLN palsies, so the RLN palsy rate according to "nerves at risk" was 963/34638 = 2.8%.

Pearson's correlation co-efficient was calculated for the predictive factors examined (patient gender, thyroid status at presentation, goitre type, routine laryngoscopy, reoperation, surgeon volume, extent of surgery, use of RLN monitor, use of additional energy device in addition to monopolar / bipolar diathermy and thyroid cancer). All these predictive factors except the use of additional energy devices in addition to monopolar and bipolar diathermy were significantly correlated ($p < 0.01$) with RLN palsy.

In binary logistic regression analysis: age, retrosternal goitre, routine laryngoscopy, re-operation, bilateral thyroidectomy and

Table 3 Probability (median, upper and lower quartile risks) of permanent hypoparathyroidism following bilateral thyroidectomy with or without lymph node dissection after categorisation of surgeon volume annual rate and taking into account clustering of cases with mixed effect logistic regression analysis

Quartile Risk	0–25 cases/year	25–50 cases/year	50–75 cases/year	75–100 cases/year	>100 cases/year
No node dissection					
Lower	4.0%	4.4%	3.0%	2.8%	2.0%
Median	6.0%	6.6%	4.5%	4.1%	3.0%
Upper	8.9%	9.8%	6.7%	6.2%	4.6%
Node dissection					
Lower	10.4%	11.5%	8.0%	7.3%	5.4%
Median	15.1%	16.6%	11.7%	10.8%	8.0%
Upper	21.4%	23.3%	16.8%	15.6%	11.8%

Table 4 Median, upper and lower quartiles values for the probability of permanent hypoparathyroidism data being missing in UKRETS over study period 1st September 2010 to 31st August 2016 across annual surgical rate category groups

Annual rate (cases/year)	Lower quartile	Median	Upper quartile
0–25	15.0%	17.0%	19.1%
25–50	15.4%	17.4%	19.5%
50–75	14.1%	15.9%	18.0%
75–100	18.8%	21.1%	23.7%
>100	14.9%	16.8%	18.9%

node dissection were significantly associated ($p < 0.01$) with an increased risk of RLN palsy. Surgeon volume and RLN monitor were significantly associated ($p < 0.01$) with a reduced risk of RLN palsy. Male gender and hyperthyroidism at presentation were not significantly associated with RLN palsy (Table 5).

The probability of RLN palsy in the surgeon AR categories, determined by GAMs analysis of PH, shows that the RLN palsy rate fell with increasing surgeon AR from a median of 4.4% in the group performing <50 cases/year, to 2.9% in surgeons undertaking 50–100 cases/year and 2.6% in the >100 cases/year group (Table 6). When surgeons are split into 25 cases/year categories there was an increase in the median RLN palsy rate between surgeons undertaking <25 to 25–50 and 50–75 to 75–100 cases/year, but a linear decline in RLN palsy when categorizing by 50 cases/year (Table 7).

Reoperation for post-operative haematoma occurred overall in 276/25038 (1.1%). Analysis of risk factors for reoperation for post-operative bleeding in 25,038 thyroidectomies showed that male gender, patient age, hyperthyroidism at presentation, retrosternal goitre and bilateral thyroidectomy were significant correlated with re-operation for post-operative bleeding, but surgeon volume, re-operative surgery, node dissection, use of additional energy sources in addition to diathermy, and thyroid cancer were not significantly correlated with this outcome. This analysis was not taken further given the absence of correlation between surgeon volume and risk of re-operation for bleeding.

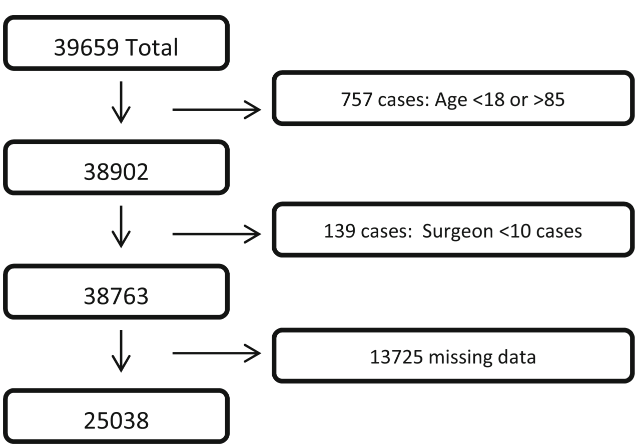

Fig. 4 Number of thyroidectomy operations analysed for probability of RLN palsy & post-operative bleeding following application of exclusion criteria

Discussion

This study set out to establish a method for categorizing surgeon outcome groups to study the association between surgeon volume and outcome with the hope of establishing a better evidence base for recommending a minimum number of thyroidectomies per year to maintain competency. It is clear from the analysis of UKRETS that there is an association between surgeon volume and outcome, and the best outcomes are occurring in surgeons performing >100 cases/year. Outcomes improved with increasing surgeon volume when data were categorized into groups of 50 cases / year. PH and RLN palsy did not decline until >50 cases/year were undertaken, suggesting 50 thyroidectomies should be the recommended minimum to maintain competence.

There are limitations to the outcome measures in UKRETS dataset investigated. PH is an important complication of thyroidectomy and is associated with hypocalcaemic symptoms, seizures, renal calcification, impaired quality of life and increased mortality [18, 19]. It is caused by surgical damage to the parathyroid glands during thyroidectomy, either via inadvertent removal, damage to vascular supply or direct trauma to the glands. It is therefore a good quality indicator of thyroid surgery. The analysis of PH was restricted to bilateral surgery as this is associated with an increased risk of hypocalcaemia, which was not always the case in previous studies.

The true incidence of RLN palsy following thyroidectomy is not known unless post-operative laryngoscopy is performed. This isn't always the case in British thyroid surgical practice. Analysis of RLN palsy in investigating volume outcome is limited for this reason. The incidence of reoperation for haematoma following thyroidectomy is low (1.1%) and this also makes it an unsuitable outcome measure when examining a volume-outcome relationship. Nevertheless RLN palsy and postoperative haematoma were included in this analysis as they are important complications of thyroid surgery.

Parathyroid glands may recover function with time following surgical damage. 63/80 (78%) with hypoparathyroidism at one month following thyroidectomy recovered parathyroid function by one year in a study by Sitges-Serra et al. [20]. It is not known what proportion of those patients with PH at 6 months follow-up in the current study went on to recover parathyroid function, as parathyroid function at 1 year is not recorded in UKRETS, though it is likely the majority did not.

This study found no association between patient gender or age and PH after adjusting for clustering of outcome of cases

Table 5 Binary logistic regression analysis of risk factors for RLN palsy following thyroid surgery

Predictive factor	Odds ratio RLN palsy	95% Confidence Interval (lower)	95% Confidence Interval (higher)	Significance (*p* value)
Male gender	1.026	0.876	1.202	ns
Hyperthyroid	0.914	0.723	1.155	ns
RLN monitor	0.662	0.576	0.761	0.000
Surgeon volume	0.916	0.894	0.937	0.000
Age	1.014	1.009	1.018	0.000
Retrosternal goitre	1.627	1.386	1.909	0.000
Bilateral thyroidectomy	1.898	1.633	2.206	0.000
Reoperation	2.038	1.678	2.475	0.000
Node dissection	2.367	1.991	2.813	0.000
Routine Laryngoscopy	9.816	8.066	11.95	0.000

from the same surgeon. Nor is there a convincing mechanism to explain an association between patient age / gender and PH.

Nodal dissection was found to be associated with PH as has been reported previously [21]. Patients undergoing a formal nodal dissection for thyroid cancer were analysed in the "nodal dissection" group in the current study. Of these >85% underwent a dissection of the central compartment lymph nodes so potentially compromising the parathyroid glands. It is unlikely that restricting the analysis to patients only undergoing central compartment dissection would have influenced the findings.

The incidence of PH reported in this study of 3% in the very high volume surgeons (>100 cases/year) is comparable to results from other high-volume centres [22, 23]. The higher incidence of PH in lower surgical volume groups is probably mirrored globally, as there is a publication bias favouring good outcomes from high volume centres. It is worth noting that UKRETS does not record data on the severity of PH, just whether vitamin D or calcium supplements are being taken to maintain normocalcaemia.

There is some cause for concern in this study, with approximately one in ten and one quarter of patients suffering PH in the highest quartile of intermediate volume surgeons (25–50 cases/year) after bilateral thyroidectomy without and with nodal dissection respectively (Table 3). The peak incidence of PH was seen in the intermediate (25–50 cases/year) volume group. It is interesting to note that the only other study from the UK based on HES data also found a peak in incidence of medical complications, vocal cord paralysis and death / life-threatening complications among surgeons performing 20–29 thyroidectomies / year compared to 10–19 per year or higher volume groups [13]. The cause for this is not known, though it is possible that other risk factors for PH such as goitre size or type (retrosternal versus cervical goitre) may have influenced the outcome.

Reported RLN palsy rates following thyroidectomy vary between 1.5–11% [11, 24–28], though recovery may occur in some cases and so the incidence of permanent RLN palsy is lower. There are a number of risk factors associated with RLN palsy including central neck dissection, advanced malignancy, age, male gender, co-morbidities, and type of operation [11, 27].

Table 6 Probability of RLN palsy (median, upper and lower quartile risks) following thyroidectomy. Total cases and those undergoing routine laryngoscopy after categorisation of surgeon annual volume rate into groups of 50 cases

Quartile Risk	<50 cases/year	50–100 cases/year	>100 cases/year
Total (n = 25038)			
Lower	0.7%	0.7%	2.0%
Median	4.4%	2.9%	2.6%
Upper	8.1%	5.9%	3.5%
Routine Laryngoscopy (*n* = 11491)			
Lower	3.5%	3.2%	2.2%
Median	6.25%	5.9%	3.4%
Upper	10.2%	6.7%	3.5%

Table 7 Probability of RLN palsy (median, upper and lower quartile risks) following thyroidectomy. Total cases and those undergoing routine laryngoscopy after categorisation of surgeon annual volume rate into groups of 25 cases

Cases/Year	<25	25-50	50-75	75-100	>100
Total (*n* = 25038)					
Lower	0%	0.9%	0.7%	3.0%	2.0%
Median	4.3%	4.4%	2.4%	4.2%	2.6%
Upper	7.5%	8.1%	4.5%	6.7%	3.5%
Routine Laryngoscopy (n = 11491)					
Lower	3.2%	3.8%	2.9%	4.2%	2.2%
Median	6.25%	5.9%	5.1%	6.2%	3.4%
Upper	9.4%	10.9%	6.3%	7.7%	3.5%

High-volume surgeons have a reduced incidence of RLN palsy and one study has reported that volumes >30 are protective [11]. In this study we've identified a number of factors associated with RLN palsy, of which use of routine (pre and post-operative) laryngoscopy appears to have the greatest influence with an odds ratio of 9.8 for the occurrence of this complication. This suggests that the incidence of RLN palsy would be higher if all cases underwent routine laryngoscopy.

There is undoubtedly an inherent bias in UKRETS data when using RLN palsy as an outcome measure to investigate a surgeon volume association, as BAETS surgeons fall largely into 2 groups - those who perform laryngoscopy selectively or non-selectively. The true incidence of RLN palsy is not known without post-op laryngoscopy, so the cases performed by "selective laryngoscopy" surgeons, who did not undergo post-operative laryngoscopy, may have included some cases with RLN palsy, that were analysed in the "no RLN palsy group". For this reason the probability of RLN palsy in patients not undergoing routine laryngoscopy in this study will be underestimated.

The pattern of declining incidence of RLN palsy with increasing surgeon volume was seen, albeit with a higher incidence of RLN palsy, in those cases that had undergone routine (i.e. pre- and post-op) laryngoscopy – see Table 6. This group included high risk cases for RLN palsy from the "selective laryngoscopy" surgeons, as well as all cases performed by "non-selective laryngoscopy" surgeons. The incidence of RLN palsy in patients undergoing routine laryngoscopy in this study is likely to be an overestimate of the overall incidence of RLN palsy following thyroidectomy.

The probability of RLN palsy in patients undergoing retrosternal goitre, bilateral thyroidectomy, re-operation and nodal dissection all significantly increased the risk of RLN palsy with an odds ratio more than 1.5. It is interesting to note that use of RLN monitoring reduced the odds ratio of RLN palsy by one third in this study.

Although increasing surgeon volume was significantly associated with reduced RLN palsy, it is difficult to interpret the clinical significance of this, given the limitations outlined above. As with PH the relationship between surgeon volume and RLN palsy is not linear when data was categorized into groups of 25 cases/year, though when data were categorized into groups of 50 cases / year the incidence of RLN palsy fell with increasing surgeon AR.

The incidence of re-operation for post-operative bleeding was low and so unsuitable for the analysis of surgeon volume outcome, despite the high number of operations analysed. Post-operative haematoma may also be a more random event, less related to surgical experience than PH or RLN palsy. It was noted that the probability of post-operative haematoma was 1.3% in surgeons performing <50 cases / year, 1.0% in 50–100 cases / year and 0.9% in the >100 cases /year group. But in the absence of any significant association between surgeon AR and haematoma it is difficult to draw further conclusions from this.

The main aim of this study was to describe a method for categorizing data for the analysis of volume outcome and examine the evidence base for recommending a minimum number of 20 thyroidectomies per year to maintain competency in the UK. Only one study in the literature has addressed this particular issue, finding a surgeon volume threshold of 25 cases / year was associated with better outcomes [12]. It is arguable if the findings of this study from North America apply to UK thyroid practice, as the data came largely from low volume surgeons – 51% of whom did an average of one case/year. In comparison the data in UKRETS has a higher proportion of intermediate 25–50 cases/year (37%) and high volume > 50 cases/year (27%) performing thyroid surgery, perhaps reflecting membership of a specialist organization dedicated to thyroid and endocrine surgery.

In this study we used a method to categorize variables that is reproducible and results in distinct, interpretable surgical volume groups. GAMs first described by Hastie and Tibshirami [29] were used to categorize the surgeon volume groups. GAMs is a method for investigating a smooth non-linear relationship between dependent and predictive variables. The advantage is that this may capture patterns missed by classic linear modelling, and the smooth non-linear curve is readily interpretable for the purposes of categorizing data. Mixed effects logistic regression was then used in analysis of PH as it allows for dependence of outcome among cases operated on by the same surgeon to be taken into account.

Registry data compiled by the clinicians, and therefore more likely to be clinically accurate than hospital administration data, was used in this study. This is subject to limitations too, highlighted by the high incidence of missing data at long-term follow up. To what extent this influenced the findings of the study are not known, as no consistent pattern was observed in missing data rates across volume groups. It does limit the reliability of and the confidence in which conclusions about UK thyroid practice can be drawn from the results.

Registry data is also subject to human data entry error which could affect results. A number of errors were found in UKRETS for LOS and at extremes of patient age which precluded the use of LOS as an outcome measure in this study. In addition, the effect of culminated surgical experience on outcome independent of volume of surgery performed during this study was not accounted for, although there is evidence that this affects outcomes from thyroid surgery [30].

Conclusions

The results of this study confirm findings of previous studies that show better outcomes from thyroid surgery in high volume surgeons. Both PH and RLN palsy following thyroidectomy were associated with surgeon volume and declined in incidence in surgeons performing >50 cases/year. The results are limited by the high missing data rate in the database, but do tentatively suggest that the threshold for minimum number of thyroidectomies needed to maintain proficiency should be 50 cases / year. Though the best outcomes occurred, in fact, in surgeons undertaking >100 cases/year.

Results of this study have been presented and discussed at the BAETS annual meeting. Any recommendation for a minimum number of cases has to take into to account the feasibility of delivering this service as the majority (73%) of BAETS members who contributed to this database in this study performed <50 cases / year and increasing specialization in thyroid surgery would inevitably result in patients travelling further distances to access surgical care.

Acknowledgements The authors thank Professor John Matthews (Newcastle University, School of Mathematics & Statistics) for his help with the statistical analysis and interpretation of PH from the UKRETS data. The authors also thank the BAETS members and those who maintain and contribute to UKRETS including Dendrite clinical systems Limited.

Contributors to UKRETS Richard Adamson, Anna Aertssen, Ahmed Afzaal, Avi Agrawal, Afzaal Ahmad, Ijaz Ahmad, Orabi Ahmad, Ibrahim Ahmed, Irfan Akhtar, Murat Akyol, Peyman Alam, Munther Aldoori, David Allen, Iain D Anderson, Sebastian Aspinall, Chris Ayshford, Ekambaram Dinakara Babu, Christopher Backhouse, Saba alasubramanian, Alistair Balfour, Neal Banga, Ludger Barthelmes, Nigel Beasley, Chris Bem, Ian Black, Stephen Blair, Richard Bliss, Victoria Brown, Robert Carpenter, Michael Carr, AndreW Carswell, Carmen de Casso Moxo, David Chadwick, Habib Charfare, Andy Chin, Edward Chisholm, Louise Clark, Peter Clarke, Helen Cocks, Peter Conboy, Luke Condon, Rogan Corbridge, Allan Paul Corder, Paul Counter, Stephen P Courtney, Eamonn Coveney, Hugh Cox, Wendy Craig, James N Crinnion, David Cunliffe, Titus Cvasciuc, Jeremy P Davis, Stuart Denholm, Ganapathy Dhanasekar, Vikram Dhar, Ann Dingle, James Docherty, Helen Doran, Julie Dunn, Fiona Eatock, Anusha Edwards, Wael Elsaify, James England, Abigail A Evans, Roy Farrell, Brian Fish, Bence Forgacs, Clare Fowler, Georgios Fragkiadakis, Gabriele Galata, Ashu Gandhi, Richard Garth, Ajith George, Nicholas Gibbins, Martin G Greaney, Thomas G Groot-Wassink, Paul Gurr, Andrew Guy, Wayne Halfpenny, Charles Hall, Paul Hans, Robert Hardy, Churunal Hari, Barney Harrison, Micheal Harron, Simon Hickey, Omar Hilmi, Tim Hoare, Jonathan Hobson, Philip Holland, Andrew Houghton, David Howe, Jonathan Hubbard, Neil Hulton, Paul Hurley, Andrew Husband, Aidah Isa, Shaun Jackson, Tony Jacob, Sharan Chakkyath Jayaram, Jean- Pierre Jeannon, Taleb Jeddy, Stephanie Jenkins, Bethan Jones, Anton Joseph, Bengt Kald, Robert Kennedy, James Kirkby-Bott Paul Kirkland, Ursula Kirkpatrick, Zygmunt Krukowski, Nirmal Kumar, Vijay Kurup, Tom Kurzawinski, Nicholas RF Lagattolla, Mark Lansdown, Nicholas Law, Tom WJ Lennard, Peter Lewis, Andrew P Locker, John RC Logie, Sean Loughran, M Lucarotti, John Lynn, Alasdair Mace, Fiona MacGregor, Paul R Maddox, Arcot Maheshwar, Zvoru Makura, Deborah Markham, Dominic Martin-Hirsch, Andrew Mccombe, Julian McGlashan, Andrew McIrvine, Andrew J McLaren, Sandy McPherson, Hesham Mehanna, Radu Mihai, Faisal MA Mihaimeed, Tedla Miroslav, George Mochloulis, James Moor, Peter Moore, Ram Moorthy, Pradeep Morar, Justin Morgan, Iain M Muir, Michael L Nicholson, Stewart Nicholson, Keshav Nigam, Iain Nixon, Janet O'Connell, Olawale Olarinde, Fausto Palazzo, Michael Papesch, Neil R Parrott, Susannah Penney, Andrew Pfleiderer, Jonathan Philpott, Lisa Pitkin, Isabel Quiroga, David Ratliff, Duraisamy Ravichandran, Venkat Reddy, David Rew, Keith Rigg, Nick Roland, Alasdair Ross, Tom Rourke, Gavin T Royle, Sarwat Sadek, Greg Sadler, Mrinal Saharay, Michael, Ahmed Samy, Klaus- Martin Schulte, David Scott-Coombes, Anup Kumar Sharma, Steve Shering, Susannah Shore, John Shotton, Richard Sim, Ricard Simo, Prakash Sinha, Gunasekaran Sinnappa, Anthony Skene, James Smellie, David M Smith, Ian Smith, Simon Smith, Roy Spence, Paul Spraggs, Adam Stacey- Clear, Frank Stafford, Michael P Stearns, Michael Stechman, Paul Stimpson, Robert Sudderick, Robert Sutclffe, Peter Tassone, Taranjit Tatla, Gareth Tervit, Paul Thomas, Adrian Thompson, Steven Thrush, Paul Tierney, Augustine Titus, Neil Tolley, Mark Tomlinson, Paul Turner, Charanjeit S Ubhi, Harpreet Uppal, Srinivasan Venkat, Richard Vowles, Alison Waghorn, Jonathan C Watkinson, Gavin Watters, John Weighill, Andrew R Welch, Hugh Wheatley, Martin Wickham, Chandana Wijewardena, Adam Wild, Michael R Williams, Simon Williams, Paul Wilson, Michail Winkler, Stephen Wood, Constantinos Yiangou, Charles Zammitt.

Author's contribution Mr. Aspinall: Study conception and design, interpretation of data, Drafting of manuscript, critical revision of manuscript. Miss Oweis: Critical revision of manuscript Mr. Chadwick: Study conception and design, interpretation of data, critical revision of manuscript.

Compliance with ethical standards

Conflict of interest S. Aspinall declares that he has no conflict of interest. D Oweis declares that she has no conflict of interest. D Chadwick declares that he has no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed consent No applicable.

References

1. Chowdhury MM, Dagash H, Pierro A (2007) A systematic review of the impact of volume of surgery and specialization on patient outcome. Br J Surg 94(2):145–161. https://doi.org/10.1002/bjs.5714
2. Boudourakis LD, Wang TS, Roman SA, Desai R, Sosa JA (2009) Evolution of the surgeon-volume, patient-outcome relationship. Ann Surg 250(1):159–165. https://doi.org/10.1097/SLA.0b013e3181a77cb3
3. Stavrakis AI, Ituarte PH, Ko CY, Yeh MW (2007) Surgeon volume as a predictor of outcomes in inpatient and outpatient endocrine surgery. Surgery 142(6):887–899; discussion 887-99. https://doi.org/10.1016/j.surg.2007.09.003
4. Sosa JA, Bowman HM, Tielsch JM, Powe NR, Gordon TA, Udelsman R (1998) The importance of surgeon experience for clinical and economic outcomes from thyroidectomy. Ann Surg 228(3):320–330
5. Dralle H, Sekulla C, Haerting J, Timmermann W, Neumann HJ, Kruse E, Grond S, Muhlig HP, Richter C, Voss J, Thomusch O, Lippert H, Gastinger I, Brauckhoff M, Gimm O (2004) Risk factors of paralysis and functional outcome after recurrent laryngeal nerve monitoring in thyroid surgery. Surgery 136(6):1310–1322. https://doi.org/10.1016/j.surg.2004.07.018
6. Gourin CG, Tufano RP, Forastiere AA, Koch WM, Pawlik TM, Bristow RE (2010) Volume-based trends in thyroid surgery. Arch

Otolaryngol Head Neck Surg 136(12):1191–1198. https://doi.org/10.1001/archoto.2010.212
7. Kandil E, Noureldine SI, Abbas A, Tufano RP (2013) The impact of surgical volume on patient outcomes following thyroid surgery. Surgery 154(6):1346–1352; discussion 1352-3. https://doi.org/10.1016/j.surg.2013.04.068
8. Loyo M, Tufano RP, Gourin CG (2013) National trends in thyroid surgery and the effect of volume on short-term outcomes. Laryngoscope 123(8):2056–2063. https://doi.org/10.1002/lary.23923
9. Adkisson CD, Howell GM, McCoy KL, Armstrong MJ, Kelley ML, Stang MT, Joyce JM, Hodak SP, Carty SE, Yip L (2014) Surgeon volume and adequacy of thyroidectomy for differentiated thyroid cancer. Surgery 156(6):1453–1459; discussion 1460. https://doi.org/10.1016/j.surg.2014.08.024
10. Hauch A, Al-Qurayshi Z, Randolph G, Kandil E (2014) Total thyroidectomy is associated with increased risk of complications for low- and high-volume surgeons. Ann Surg Oncol 21(12):3844–3852. https://doi.org/10.1245/s10434-014-3846-8
11. Nouraei SA, Virk JS, Middleton SE, Aylin P, Mace A, Vaz F, Kaddour H, Darzi A, Tolley NS (2017) A national analysis of trends, outcomes and volume-outcome relationships in thyroid surgery. Clin Otolaryngol 42(2):354–365. https://doi.org/10.1111/coa.12730
12. Adam MA, Thomas S, Youngwirth L, Hyslop T, Reed SD, Scheri RP, Roman SA, Sosa JA (2017) Is there a minimum number of thyroidectomies a surgeon should perform to optimize patient outcomes? Ann Surg 265(2):402–407. https://doi.org/10.1097/SLA.0000000000001688
13. Nouraei SA, Hudovsky A, Frampton AE, Mufti U, White NB, Wathen CG, Sandhu GS, Darzi A (2015) A study of clinical coding accuracy in surgery: implications for the use of administrative big data for outcomes management. Ann Surg 261(6):1096–1107. https://doi.org/10.1097/SLA.0000000000000851
14. Chadwick D, Kinsman R, Walton P (2017) Fifth National audit report, British Association of Endocrine and Thyroid Surgeons, Henley-on-Thames : Dendrite Clinical Systems
15. Hospital Episode Statistics. NHS Digital. In: https://digital.nhs.uk/data-and-information/data-tools-and-services/data-services/hospital-episode-statistics
16. National Cancer Peer Review (2013) Manual for cancer services. Head and Neck Measures. National Cancer Action Team, London
17. Chadwick D, Kinsman R, Walton P (2012) Fourth national audit report. British Association of Endocrine and Thyroid Surgeons, Henley-on
18. Mitchell DM, Regan S, Cooley MR, Lauter KB, Vrla MC, Becker CB, Burnett-Bowie SA, Mannstadt M (2012) Long-term follow-up of patients with hypoparathyroidism. J Clin Endocrinol Metab 97(12):4507–4514. https://doi.org/10.1210/jc.2012-1808
19. Almquist M, Ivarsson K, Nordenstrom E, Bergenfelz A (2018) Mortality in patients with permanent hypoparathyroidism after total thyroidectomy. Br J Surg 105(10):1313–1318. https://doi.org/10.1002/bjs.10843
20. Sitges-Serra A, Ruiz S, Girvent M, Manjon H, Duenas JP, Sancho JJ (2010) Outcome of protracted hypoparathyroidism after total thyroidectomy. Br J Surg 97(11):1687–1695. https://doi.org/10.1002/bjs.7219
21. Roh JL, Kim JM, Park CI (2011) Central lymph node metastasis of unilateral papillary thyroid carcinoma: patterns and factors predictive of nodal metastasis, morbidity, and recurrence. Ann Surg Oncol 18(8):2245–2250. https://doi.org/10.1245/s10434-011-1600-z
22. Asari R, Passler C, Kaczirek K, Scheuba C, Niederle B (2008) Hypoparathyroidism after total thyroidectomy: a prospective study. Arch Surg 143(2):132–137; discussion 138. https://doi.org/10.1001/archsurg.2007.55
23. Almquist M, Hallgrimsson P, Nordenström E, Bergenfelz A (2014) Prediction of permanent hypoparathyroidism after total thyroidectomy. World J Surg 38(10):2613–2620. https://doi.org/10.1007/s00268-014-2622-z
24. Serpell JW, Lee JC, Yeung MJ, Grodski S, Johnson W, Bailey M (2014) Differential recurrent laryngeal nerve palsy rates after thyroidectomy. Surgery 156(5):1157–1166. https://doi.org/10.1016/j.surg.2014.07.018
25. Thomusch O, Sekulla C, Walls G, Machens A, Dralle H (2002) Intraoperative neuromonitoring of surgery for benign goiter. Am J Surg 183(6):673–678
26. Yarbrough DE, Thompson GB, Kasperbauer JL, Harper CM, Grant CS (2004) Intraoperative electromyographic monitoring of the recurrent laryngeal nerve in reoperative thyroid and parathyroid surgery. Surgery 136(6):1107–1115. https://doi.org/10.1016/j.surg.2004.06.040
27. Dhillon VK, Rettig E, Noureldine SI, Genther DJ, Hassoon A, Al Khadem MG, Ozgursoy OB, Tufano RP (2018) The incidence of vocal fold motion impairment after primary thyroid and parathyroid surgery for a single high-volume academic surgeon determined by pre- and immediate post-operative fiberoptic laryngoscopy. Int J Surg 56:73–78. https://doi.org/10.1016/j.ijsu.2018.06.014
28. Joliat GR, Guarnero V, Demartines N, Schweizer V, Matter M (2017) Recurrent laryngeal nerve injury after thyroid and parathyroid surgery: incidence and postoperative evolution assessment. Medicine (Baltimore) 96(17):e6674. https://doi.org/10.1097/MD.0000000000006674
29. Hastie T, Tibshirani R (1986) Generalized additive models. Stat Sci 1(3):297–318
30. Duclos A, Peix JL, Colin C, Kraimps JL, Menegaux F, Pattou F, Sebag F, Touzet S, Bourdy S, Voirin N, Lifante JC, Cathy Study Group (2012) Influence of experience on performance of individual surgeons in thyroid surgery: prospective cross sectional multicentre study. BMJ 344:d8041. https://doi.org/10.1136/bmj