



# The volume and outcome relationship for thyroidectomy in England

William K. Gray<sup>1</sup> · Sebastian Aspinall<sup>2,3</sup> · Neil Tolley<sup>4</sup> · Jamie Day<sup>1</sup> · Mark Lansdown<sup>1,5</sup>

Received: 1 July 2020 / Accepted: 1 June 2021 / Published online: 9 June 2021

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

## Abstract

**Purpose** The delivery of surgical care in England has seen a momentum towards centralisation within larger volume hospitals and surgical teams. The aim of this study was to investigate outcomes in England in relationship to hospital and surgeon annual volumes for total thyroidectomy.

**Methods** Data were extracted from the Hospital Episodes Statistics (HES) database for England. A 6-year period (April 2012–March 2018 inclusive) for all adult admissions for thyroidectomy was used in the analysis. The primary outcome measure used was a length of hospital stay greater than 2 days or an emergency readmission within 30 days following surgery. This was used as a proxy for surgical complications. A multilevel modelling strategy was used to adjust for hierarchy and potentially confounding.

**Results** Data for 22,823 total thyroidectomies across 144 hospital trusts were used for analysis. For total thyroidectomy, larger volume surgeons had reduced levels of post-surgical complications; length of stay > 2 and > 4 days; emergency readmission at 30 days; and hypoparathyroidism, vocal cord palsy, stridor, and tracheostomy at 1-year post-surgery. Larger hospital volume was associated with lower levels of emergency readmission at 30 days and hypoparathyroidism at 1 year.

**Conclusions** There is significant correlation between surgeon volume and clinical outcome for total thyroidectomy. The relationship was approximately linear, and a low-volume threshold could not be defined.

**Keywords** Thyroidectomy · Thyroid surgery · Volume-outcome relationship · Hypoparathyroidism · Vocal cord palsy · Tracheostomy

## Introduction

The devastating financial impact of the coronavirus pandemic will herald significant change in how healthcare will be delivered in the future. Economies of scale will undoubtedly take place with an increasing focus on a hub-spoke model of care. An emphasis on quality of care will still be

at the vanguard, with patient outcomes being a driver for change.

Currently, the National Health Service (NHS) in England has a focus on the consolidation and delivery of surgical care by creating a framework of excellence in delivery, research, training, and education. The concentration of specified operations within high-volume centres and surgeons would be attractive in achieving this aim [1, 2].

There is abundant extant literature supporting the premise that larger surgical volumes lead to more favourable clinical outcomes, which benefits patients and the healthcare economy [3]. However, reorganisation is by its nature, disruptive, not without cost and may negatively impact on patient service access, training and the sustainability of departments and smaller hospitals [4].

The Getting It Right First Time (GIRFT) programme, funded by the Department of Health and Social Care in England, has a remit to reduce unwarranted variation in healthcare provision. The GIRFT programme has studied the impact on patients and healthcare resources of low-volume

✉ William K. Gray  
William.gray5@nhs.net

<sup>1</sup> Getting It Right First Time Programme, NHS England and NHS Improvement, 1 Lower Marsh, London SE1 7NT, UK

<sup>2</sup> Aberdeen Royal Infirmary, NHS Grampian, Aberdeen, UK

<sup>3</sup> Audit Lead, British Association of Endocrine and Thyroid Surgeons, London, UK

<sup>4</sup> Hammersmith Hospital, Imperial College Healthcare NHS Trust, London, UK

<sup>5</sup> St. James's University Hospital, Leeds Teaching Hospitals NHS Trust, Leeds, UK

surgery. It has sought to provide evidence as to whether patient outcomes are linked to surgical volume to better inform national decision-makers.

For thyroid surgery in England, Nouraei et al. [5], analysing Hospital Episode Statistics (HES) data over 6 years, identified fewer medical and surgical complications, shorter length of stay, and lower rates of vocal cord palsy as associated with surgeons with larger annual volumes of thyroidectomy. Greater than 30 procedures per annum consistently showed to be protective. Aspinnall et al. [6] in 2019 published data on 25,038 thyroidectomies from the United Kingdom Registry of Endocrine and Thyroid Surgery (UKRETS), which is organised by the British Association of Endocrine and Thyroid Surgeons (BAETS). They identified a link between surgeon volume and both permanent hypoparathyroidism and recurrent laryngeal nerve palsy. The authors proposed a minimal annual volume of 50 thyroidectomies for each surgeon. This differed from the recommendations made from the majority of previous studies on volume outcomes from the USA. Adkisson et al. [7] identified 30 as a cut-off for annual surgeon volumes with regard to outcomes for completeness of resection and the appropriateness of the operation undertaken. Adam et al. [8] identified less than 26 procedures per annum to be associated with an increased complication rate. In contrast, a recent study showed that although larger surgeon volumes were associated with lower rates of vocal cord palsy and hypoparathyroidism, the absolute effect of volume on outcomes was modest, especially in hospitals with low complication rates [9]. Although the evidence for a volume-outcome relationship in thyroidectomy appears strong, other confounding surgeon and hospital-specific factors may also play an important role [10]. Nevertheless, the American Thyroid Association and the American Association of Endocrine Surgeons both recommend that thyroidectomy should be performed by high-volume surgeons where possible, defined as a minimum of 20 thyroidectomies annually [11, 12].

The aim of this study was to use the HES database to gain insight into the nature and extent of the volume-outcome relationship for thyroidectomy in England. To increase the homogeneity within our dataset, we used total thyroidectomy as an exemplar condition.

## Methods

### Study design

A retrospective analysis of data from HES. HES data are collected by NHS Digital for all patients treated at English hospitals that are funded by the NHS. In England, health-care coverage is universal, free at the point of delivery, and funded from general taxation. As such, all but a small

minority of thyroid surgery is conducted within the NHS system, with very few patients seen by private providers.

To investigate volume-outcome relationships, a relatively homogenous dataset with regard to procedure type is required. Total thyroidectomy was chosen as an exemplar condition. Total thyroidectomy is performed in sufficiently large volumes nationally to facilitate a meaningful analysis, and outcomes are not impacted, in general, by previous thyroid surgery. Furthermore, complications for total thyroidectomy are greater than those of partial thyroidectomy [13].

### Ethics

Consent from individuals involved in this study was not required. The analysis and presentation of data follows current NHS Digital guidance for the use of HES data for research purposes and is anonymised to the level required by ISB1523 Anonymisation Standard for Publishing Health and Social Care Data [14, 15].

### Case identification in HES

*Setting:* hospital admissions in England.

*Timing:* financial years 2012–2013 to 2017–2018.

*Admission method:* elective or day-case.

*Age:*  $\geq 16$  years.

*Diagnosis:* The Classification of Interventions and Procedures version 4 (OPCS-4) codes listed in Table 1 were used to identify thyroidectomy procedures.

*Exclusions from modelling of volume-outcome relationships in total thyroidectomy:* Laryngectomy for head and neck cancer was excluded from the analysis. The OPCS codes used to identify laryngectomy and the International Statistical Classification of Disease and Related Health Problems 10th revision (ICD-10) codes used to identify patients with a diagnosis of head and neck cancer are detailed in Table 1.

For the purpose of statistical modelling, where a patient had multiple admissions, only the chronologically first admission was retained. This ensured that all admissions were independent of one another at a patient level. See the “[Data management and statistical analyses](#)” section for more details.

### Data extraction

*Exposure:* volume of surgery per trust and per surgeon per annum.

*Primary outcome:* emergency hospital readmission within 30 days or length of hospital stay  $> 2$  days, or both. This primary outcome was chosen as it was felt to be the best reflection of a complication of surgery, with a need for either an extended hospital stay or an emergency readmission soon

**Table 1** Diagnostic and procedural codes used to identify thyroidectomy and subsequent complication

<b>Thyroidectomy</b>	<b>Procedure</b>	<b>OPCS procedural code</b>
	Total thyroidectomy	B081
	Subtotal thyroidectomy	B082
	Hemithyroidectomy	B083
	Lobectomy of thyroid gland	B084
	Isthmectomy of thyroid gland	B085
	Partial thyroidectomy	B086
	Other specified excision of thyroid gland	B088
	Unspecified excision of thyroid gland	B089
	Diagnosis of head or neck cancer	C01X, C051, C052, C07X, C080, C081, C089, C090, C091, C098, C099, C100, C101, C102, C103, C108, C109, C110, C111, C112, C113, C118, C119, C12X, C130, C131, C132, C138, C139, C320, C321, C322, C328, C329
<b>Head and neck cancer</b>	<b>Diagnosis</b>	<b>ICD-10 diagnostic code</b>
	Diagnosis of head or neck cancer	C01X, C051, C052, C07X, C080, C081, C089, C090, C091, C098, C099, C100, C101, C102, C103, C108, C109, C110, C111, C112, C113, C118, C119, C12X, C130, C131, C132, C138, C139, C320, C321, C322, C328, C329
<b>Thyroid cancer</b>	<b>Diagnosis</b>	<b>ICD-10 diagnostic code</b>
	Malignant neoplasm of thyroid gland	C73
<b>Laryngectomy</b>	<b>Procedure</b>	<b>OPCS procedural code</b>
	Total laryngectomy	E291
	Partial horizontal laryngectomy	E292
	Partial vertical laryngectomy	E293
	Other partial laryngectomy	E294
	Laryngofissure and chordectomy of vocal cord	E295
	Other laryngectomy	E296
	Other specified excision of larynx	E298
	Unspecified excision of larynx	E299
<b>Hypoparathyroidism/ calcium disorder</b>	<b>Diagnosis</b>	<b>ICD-10 diagnostic code</b>
	Post-procedural hypoparathyroidism	E892
	Disorders of calcium metabolism	E835
	Tetany	R290
<b>Vocal cord palsy</b>	<b>Diagnosis</b>	<b>ICD-10 diagnostic code</b>

**Table 1** (continued)

	Paralysis of vocal cords and larynx	J380
	Stenosis of larynx	J386
	<b>Procedure</b>	<b>OPCS procedural code</b>
	Repair of vagus nerve	A307
	Repair of specified cranial nerve	A308
	Laryngotracheal reconstruction using cartilage graft	E311
	Laryngotracheoplasty NEC	E312
	Division of stenosis of larynx and insertion of prosthesis into larynx	E313
	Other specified reconstruction of larynx	E318
	Unspecified reconstruction of larynx	E319
	External arytenoidectomy	E331
	Cordopexy of vocal cord	E332
	Operations on cartilage of larynx NEC	E333
	Vocal cord medialisation using implant	E335
	Vocal cord medialisation using biological material	E336
	Endoscopic vocal cord medialisation	E357
	Microtherapeutic endoscopic extirpation of lesion of larynx using laser	E341
	Endoscopic arytenoidectomy	E351
	Injection into larynx	E381
<b>Tracheostomy</b>	<b>Procedure</b>	<b>OPCS procedural code</b>
	Permanent tracheostomy	E421
	Temporary tracheostomy	E423
<b>Stridor</b>	<b>Diagnosis</b>	<b>ICD-10 diagnostic code</b>
	Stridor	R061

after surgery. The choice of 2 days was based on the experience and judgement of the authors but was supported by the data (see the “**Results**” section).

*Secondary outcomes:* vocal cord palsy within 1 year, post-procedural hypoparathyroidism within 1 year, stridor diagnosis within 1 year, tracheostomy (either permanent or temporary) within 1 year and length of stay > 4 days. In addition, length of stay > 2 days and emergency hospital readmission within 30 days were considered independent outcomes. The codes detailed in Table 1 are used to identify subsequent hospital admissions where the specified diagnosis was present or where the specified procedure was undertaken. These codes are based on those used by Nouraei et al. [5]. Where a patient had multiple admissions with the same code, only the first instance where the code was used was retained. This admission was then used to calculate the time to the complication being present.

*Covariates:* age, sex, Hospital Frailty Risk Score (HFRS) band [16], financial year of admission, day-case

surgery (not used for outcomes involving length of stay), diagnosis of thyroid cancer, and surgeon speciality (ear, nose and throat, general surgery, other).

*Other data extracted:* all-cause mortality for 1-year follow-up from the UK Office for National Statistics (ONS).

### Data management and statistical analyses

Data were extracted onto a secure encrypted Structured Query Language (SQL) server controlled by NHS England and NHS Improvement. Analysis within this secure environment took place using standard statistical software: Microsoft Excel (Microsoft Corp, Redmond, WA, USA), SAS (SAS Institute Inc., Cary, NC, USA), and Alteryx (Alteryx Inc., Irvine, CA, USA). Standard descriptive statistics (e.g. mean, median, frequency) are used as appropriate to the level of the data. Age was broadly normally distributed.

## Calculation of annual volumes

Annual hospital and surgeon volume were calculated by taking the total volume of all thyroidectomy procedures across the 6-year study period and dividing it by the total number of years the trust or surgeon contributed data. For exploratory descriptive analysis and identification of a possible volume threshold, volume data were categorised into groups of 1–4, 5–9, 10–15 etc. per annum up to 45–49 then 50–59, 60–69, 70–79 etc. per annum for surgeon volume and 1–19, 20–39, 40–59 etc. per annum for hospital volume. Larger volume categories were combined to ensure there were at least five hospital trusts and surgeons represented in each volume category.

## Modelling of factors associated with outcomes for total thyroidectomy

Initially, the relationship between volume and the primary outcome was modelled using restricted cubic splines to investigate whether the relationship between volume and the log odds of the outcome was linear. Hospital volume and surgeon volume models with five equally spaced knots were developed. In neither of the two models were any of the three cubic terms significant and it was concluded that the relationships were approximately linear in both cases. Subsequently, conventional multilevel, multivariable logistic regression modelling was used for the analyses presented.

Multilevel modelling allowed for the hierarchical nature of the data and was implemented using the GLIMMIX protocol in SAS. Two-level (patients nested within trusts) intercept-only models were constructed using a logit link function. Volume and age were treated as continuous variables and the assumption of linearity to the logit checked using the Box-Tidwell test. Adjustment was made for the covariates listed above with only significant covariates retained in the final model. The results of these analyses are presented in terms of odds ratios (ORs) and confidence intervals (CIs).

## Calculation of population attributable risk

For the primary outcome measure, the population attributable risk (PAR) was calculated as the difference between the probability of the outcome in the entire population and the adjusted probability (from the multilevel models) of the outcome in those surgeons with volumes above a given value. As an example, if the probability of an outcome were 0.3 (30%) in the entire sample and 0.2 (20%) in surgeons with a volume  $\geq 60$  per annum, the PAR for surgeon volume  $< 60$  per annum would be 0.1 (10%).

The population impact number (PIN) was calculated as the reciprocal of the PAR. The PIN can be interpreted as the average number of patients in whom one adverse event

can be attributed to surgery at a below threshold volume [17]. In the example above, the PIN would be 10 (1/0.1) at volumes  $< 60$ , indicating that, for volumes below this threshold, on average, one adverse event in every ten would be attributable to surgery by a below threshold surgeon. For further details on the calculation of the PAR and PIN, see Supplementary Material.

## Role of the funding source

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

## Results

### Data for all thyroid procedures

The study flow diagram (Fig. 1) shows how data were extracted from HES. Data were available for 68,368

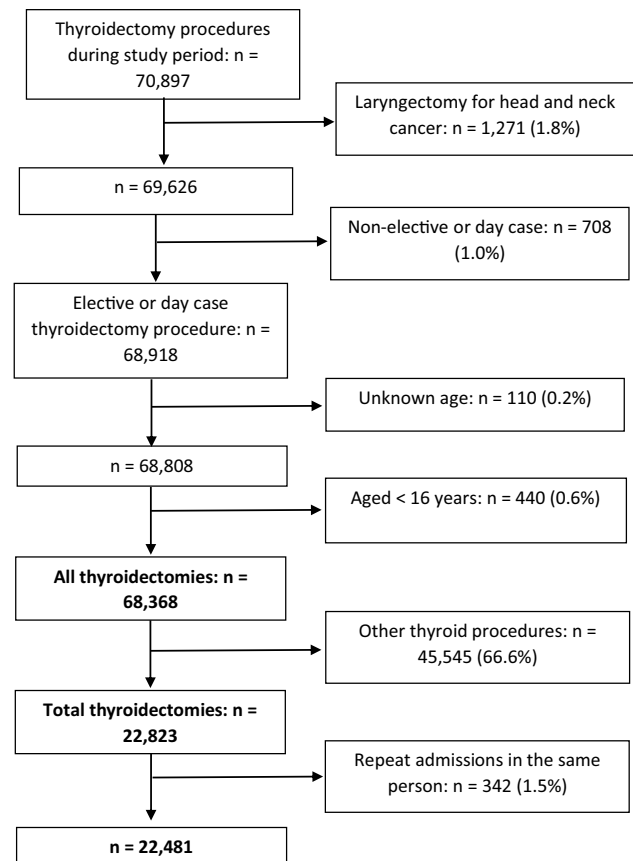


Fig. 1 Study flow diagram

thyroidectomies across 144 hospital trusts over the 6-year study period. There were 22,823 total thyroidectomies (33.4%); 41,413 lobectomies (60.6%); and 4132 other procedures (6.0%). Details of the demographic characteristics, clinical presentation, and outcomes for these procedures are shown in Table 2. Patients undergoing total thyroidectomy were slightly younger and had a longer stay, higher 30-day readmission rates, and higher rates of hypoparathyroidism at 1 year than other procedures. Those undergoing lobectomy were more likely to be seen by an ENT surgeon than those undergoing total thyroidectomy, and 3189 (7.7%) lobectomies were second lobectomy procedures in the same patient. Only 0.9% of total thyroidectomies were performed as day-cases compared to 6.9% for lobectomy. To investigate the validity of considering length of stay > 2 days as a proxy for post-operative complication, we investigated the association between hypoparathyroidism/hypocalcaemia (one of the most common immediate post-operate complications) recorded during the index admission and length of stay > 2 days. Of 1512 patients with hypoparathyroidism/hypocalcaemia recorded during the index admission, 1124 (74.3%) had a stay > 2 days.

### Changes in practice and outcomes over time for total thyroidectomy

For total thyroidectomy, details of changes in practice and outcomes over the study period are summarised in Table 3. The number of total thyroidectomies conducted per year was relatively stable with a modest fall in the number of hospital trusts performing thyroidectomy over the study period. Although the number of surgeons performing total thyroidectomy was stable, there was a clear trend towards a larger proportion of procedures being performed by ENT surgeons and fewer by general surgeons. Analysis of outcomes revealed that only length of stay showed a trend over time with a significant decline in stays longer than 2 and 4 days. There was no obvious change with regard to day-case rate over time; this was very low throughout the 6-year period.

### Volume-outcome relationships for total thyroidectomy

There was a very strong correlation between hospital (Spearman's  $r = 0.803$ ,  $p < 0.001$ ) and surgeon (Spearman's  $r = 0.669$ ,  $p < 0.001$ ) mean annual volume for total thyroidectomy and all other thyroid procedures.

Figure 2 shows the variation in the primary outcome with hospital and surgeon volume categories. On visual inspection, there is a broadly linear relationship between surgeon volume and poorer outcome although there is no clear relationship between hospital volume and outcome.

Supplementary Material Tables S1 and S2 summarise outcome data for each trust and surgeon volume category for all outcomes investigated.

Table 4 shows the findings from the multilevel logistic regression models investigating the association between volume and outcomes, after adjusting for significant covariates. A larger surgeon volume was significantly associated with lower rates of all primary and secondary outcomes investigated. A larger hospital trust volume was significantly associated with lower rates of 30-day emergency readmission and the presence of hypoparathyroidism at 1 year.

PAR and PIN values for the primary outcome measure are presented in Table 5. The PAR increased in a broadly linear fashion with increasing surgical volume with no point of inflexion at any particular volume threshold.

### Comparison of annual volumes reported in HES and UKRETS for total thyroidectomy

From HES, of the 735 surgeons performing total thyroidectomy, 256 (34.8%) had an annual volume of all thyroidectomy procedures  $\geq 20$ , 103 (14.0%) had annual volumes  $\geq 40$ , and 43 (5.9%) had annual volumes  $\geq 60$ . Within the UKRETS dataset, data were available for 41,397 thyroid procedures conducted during the same 6-year period. Of these, 13,187 (31.9%) were total thyroidectomy. Of the 243 surgeons conducting total thyroidectomy, 147 (60.5%) had an annual volume of all thyroidectomy procedures  $\geq 20$ , 74 (30.5%) had annual volumes  $\geq 40$ , and 33 (13.6%) had annual volumes  $\geq 60$ .

## Discussion

This study supports the findings of previously published literature that there is a surgeon volume-outcome relationship for total thyroidectomy performed in England [5, 6]. A relationship between hospital volume and outcomes was harder to detect and was only observed for hypoparathyroidism and 30-day emergency readmission.

The trend for improved outcomes with a larger surgeon volume was approximately linear for the primary outcome measure chosen. It was not possible to provide guidance on a minimum annual volume threshold. As such, we suggest an approach which seeks to minimise low-volume thyroid surgery as an initial priority. In this regard, we support the BAETS recommendation that in order to maintain expertise, surgeons performing thyroidectomy should perform at least 20 such procedures per year [18]. This volume is at the lower end of previous suggestions for minimum volume and may be conservative [5–9]. Regular audit and review of patient outcome data in relation to surgeon volumes, and the impact of such changes on local service provision, service viability,

**Table 2** Summary of patients' demographic, clinical characteristics, and outcomes for each procedure

	Total thyroidectomy	Subtotal thyroidectomy	Hemithyroidectomy/lobectomy of thyroid gland*	Isthmectomy of thyroid gland	Partial thyroidectomy	Other specified/unspecified thyroidectomy**
Number of patients	22,823	869	41,413	1171	838	1254
Number of trusts performing the procedure	139	110	145	119	110	123
Number of surgeons performing the procedure	735	242	837	319	299	327
Mean age (SD)	48.8 (15.870)	49.3 (15.949)	51.7 (15.450)	50.2 (14.231)	54.5 (15.873)	52.3 (16.214)
Number of females	17,988 (78.8%)	678 (78.0%)	32,726 (79.0%)	958 (81.8%)	608 (72.6%)	941 (75.0%)
HFRS band						
None	20,138 (88.2%)	751 (86.4%)	36,396 (87.9%)	1071 (91.5%)	699 (83.4%)	1062 (84.7%)
Mild	2067 (9.1%)	100 (11.5%)	3803 (9.2%)	83 (7.1%)	98 (11.7%)	146 (11.6%)
Moderate	567 (2.5%)	18 (2.1%)	1126 (2.7%)	17 (1.5%)	37 (4.4%)	45 (3.6%)
Severe	51 (0.2%)	0	88 (0.2%)	0	4 (0.5%)	1 (0.1%)
Day-case surgery present	213 (0.9%)	27 (3.1%)	2854 (6.9%)	346 (29.5%)	67 (8.0%)	57 (4.5%)
Thyroid cancer	6371 (27.9%)	149 (17.1%)	8044 (19.4%)	161 (13.7%)	192 (22.9%)	525 (41.8%)
Speciality of surgeon						
Ear, nose, and throat	11,616 (50.9%)	342 (39.4%)	27,285 (65.9%)	776 (66.3%)	538 (64.2%)	783 (62.4%)
General surgery	10,501 (46.0%)	506 (58.2%)	13,012 (31.4%)	353 (30.1%)	261 (31.1%)	413 (32.9%)
Other	706 (3.1%)	21 (2.4%)	1116 (2.7%)	42 (3.6%)	39 (4.7%)	58 (4.6%)
Length of stay > 2 days	7089 (31.1%)	209 (24.1%)	4050 (9.8%)	53 (4.5%)	173 (20.6%)	281 (22.4%)
Length of stay > 4 days	2684 (11.8%)	73 (8.4%)	1582 (3.8%)	30 (2.6%)	99 (11.8%)	117 (9.3%)
Emergency hospital readmission within 30 days	1789 (7.8%)	64 (7.4%)	1525 (3.7%)	19 (1.6%)	43 (5.1%)	69 (5.5%)
All-cause mortality within 1 year	330 (1.4%)	18 (2.1%)	355 (0.9%)	2 (0.2%)	16 (1.9%)	24 (1.9%)
Diagnosis of hypoparathyroidism within 1 year	2582 (11.3%)	64 (7.4%)	1044 (2.5%)	23 (2.0%)	48 (5.7%)	97 (7.7%)
Diagnosis of vocal cord palsy within 1 year	559 (2.4%)	16 (1.8%)	532 (1.3%)	11 (0.9%)	22 (2.6%)	26 (2.1%)
Diagnosis of stridor within 1 year	129 (0.6%)	6 (0.7%)	51 (0.1%)	2 (0.2%)	5 (0.6%)	4 (0.3%)
Temporary or permanent tracheostomy within 1 year	318 (1.4%)	11 (1.3%)	549 (1.3%)	12 (1.0%)	34 (4.1%)	30 (2.4%)
Permanent tracheostomy within 1 year	10 (0.04%)	0	17 (0.04%)	1 (0.09%)	1 (0.12%)	1 (0.08%)

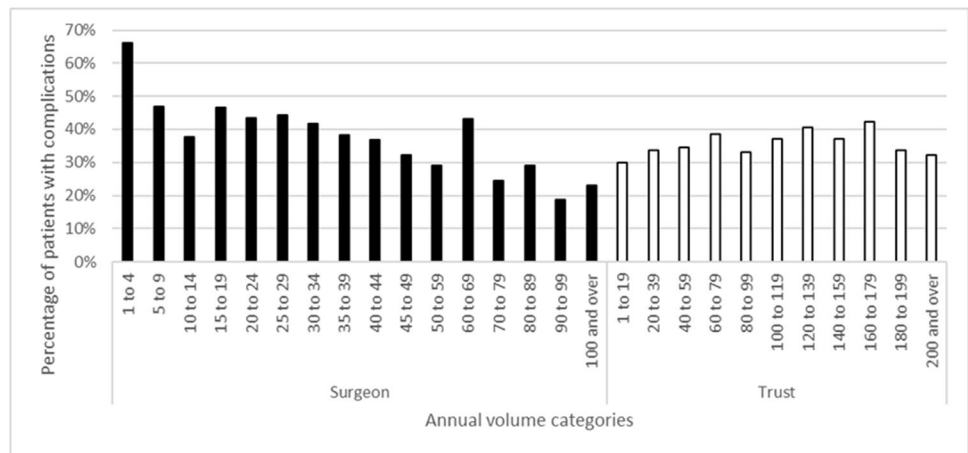
\*The OPCS-4 codes for hemithyroidectomy and lobectomy of thyroid gland were combined for these procedures

\*\*The OPCS-4 codes for other specified thyroidectomy and other unspecified thyroidectomy were combined for these procedures

**Table 3** Summary of patients' clinical characteristics and outcomes for total thyroidectomy over time

	2012–13	2013–2014	2014–2015	2015–2016	2016–2017	2017–2018
Number of patients	3722	3948	3906	3738	3835	3674
Number of trusts performing total thyroidectomy	132	128	130	128	125	123
Number of trusts performing $\geq 60$ total thyroidectomy procedures per year	14 (10.6%)	19 (14.8%)	21 (16.2%)	18 (14.1%)	15 (12.0%)	15 (12.2%)
Number of surgeons performing the procedure	390	376	377	386	393	388
Number of surgeons performing $\geq 20$ total thyroidectomy procedures per year	56 (14.4%)	52 (13.8%)	59 (15.6%)	56 (14.5%)	58 (14.8%)	57 (14.7%)
Speciality of surgeon						
Ear, nose, and throat	1653 (44.4%)	1878 (47.6%)	1991 (51.0%)	1919 (51.3%)	2127 (55.5%)	2048 (55.7%)
General surgery	1976 (53.1%)	1978 (50.1%)	1773 (45.4%)	1671 (44.7%)	1582 (41.3%)	1521 (41.4%)
Other	93 (2.5%)	92 (2.3%)	142 (3.6%)	148 (4.0%)	126 (3.3%)	105 (2.9%)
Thyroid cancer	1002 (26.9%)	1080 (27.4%)	1037 (26.5%)	1070 (28.6%)	1094 (28.5%)	1088 (29.6%)
Number of cases as day-case surgery	35 (0.9%)	31 (0.8%)	45 (1.2%)	36 (1.0%)	24 (0.6%)	42 (1.1%)
Length of stay > 2 days	1289 (34.6%)	1270 (32.2%)	1210 (31.0%)	1135 (30.4%)	1133 (29.5%)	1052 (28.6%)
Length of stay > 4 days	503 (13.5%)	482 (12.2%)	449 (11.5%)	421 (11.3%)	445 (11.6%)	384 (10.5%)
Emergency hospital readmission within 30 days	316 (8.5%)	281 (7.1%)	321 (8.2%)	272 (7.3%)	306 (8.0%)	293 (8.0%)
All-cause mortality within 1 year	72 (1.9%)	52 (1.3%)	56 (1.4%)	52 (1.4%)	49 (1.3%)	49 (1.3%)
Diagnosis of hypoparathyroidism within 1 year	442 (11.9%)	409 (10.4%)	410 (10.5%)	404 (10.8%)	445 (11.6%)	472 (12.8%)
Diagnosis of vocal cord palsy within 1 year	88 (2.4%)	79 (2.0%)	85 (2.2%)	99 (2.6%)	100 (2.6%)	108 (2.9%)
Diagnosis of stridor within 1 year	25 (0.7%)	26 (0.7%)	19 (0.5%)	21 (0.6%)	24 (0.6%)	14 (0.4%)
Temporary or permanent tracheostomy within 1 year	65 (1.7%)	47 (1.2%)	42 (1.1%)	65 (1.7%)	52 (1.4%)	47 (1.3%)
Permanent tracheostomy within 1 year	4 (0.11%)	1 (0.03%)	1 (0.03%)	1 (0.03%)	1 (0.03%)	2 (0.05%)

**Fig. 2** Variation in portion of patients with complications, defined as length of stay > 2 days or readmission within 30 days, by surgeon and trust volume categories



costs, and patient satisfaction, should help national decision-makers to make informed choices with regard to the ongoing suitability of this threshold.

The consistent finding of improved patient outcomes being linked to larger surgical volumes has been shown in many previous studies [5, 8]. Nouraei et al. [5], using HES data, discovered poorer outcomes at very low volumes, followed by a more gradual trend towards better outcomes with progressively larger volumes. A recent analysis of 16,954 patients from the US Healthcare Utilization Project

Nationwide Inpatient Sample (HCUP-NIS) dataset identified a threshold of 26 total thyroidectomies per annum as a possible threshold based on the occurrence of any complication [8]. A study using UKRETS data by Aspinall et al. [6] noted a similar trend for permanent hypoparathyroidism with a similar slight increase in complications at annual volumes above 100 as seen in this study.

Recent American Association of Endocrine Surgeons guidance made a strong recommendation that: “When possible, thyroidectomy should be performed by a high-volume



**Table 4** Adjusted odds ratios for volume measures as predictors of outcome

	Surgeon, OR, (95% CI), significance	Trust, OR, (95% CI), significance
Primary outcome		
Length of stay $\geq$ 2 days or emergency readmission to hospital within 30 days*	<b>0.989 (0.987 to 0.990), p &lt; 0.001</b>	1.000 (0.998 to 1.002), p = 0.654
Secondary outcomes		
Emergency readmission to hospital within 30 days**	<b>0.996 (0.994 to 0.998), p = 0.001</b>	<b>0.998 (0.997 to 1.000), p = 0.005</b>
Length of stay $\geq$ 2 days*	<b>0.988 (0.986 to 0.989), p &lt; 0.001</b>	1.000 (0.997 to 1.003), p = 0.993
Length of stay $\geq$ 4 days*	<b>0.988 (0.986 to 0.990), p &lt; 0.001</b>	1.002 (1.000 to 1.004), p = 0.126
Diagnosis of hypoparathyroidism within 1 year of admission***	<b>0.993 (0.991 to 0.995), p &lt; 0.001</b>	<b>0.998 (0.997 to 0.999), p = 0.004</b>
Diagnosis of vocal cord palsy within 1 year of admission****	<b>0.992 (0.989 to 0.996), p &lt; 0.001</b>	0.999 (0.997 to 1.002), p = 0.604
Diagnosis of stridor within 1 year of admission***	<b>0.991 (0.984 to 0.998), p = 0.012</b>	0.999 (0.996 to 1.002), p = 0.572
Temporary or permanent tracheostomy within 1 year****	<b>0.987 (0.981 to 0.994), p &lt; 0.001</b>	1.003 (1.000 to 1.005), p = 0.083

Notes: Based on analysis of 22,406 unique patients with a surgeon listed. Significant associations at the 5% level are shown in bold

\*Adjusted for age, sex, Hospital Frailty Risk Score band, financial year, surgeon speciality, thyroid cancer present

\*\*Adjusted for Hospital Frailty Risk Score band and surgeon speciality

\*\*\*Adjusted for age, sex, Hospital Frailty Risk Score band, surgeon speciality, and thyroid cancer present

\*\*\*\*Adjusted for age, sex, surgeon speciality, and thyroid cancer present

**Table 5** Population attributable risk and population impact number based on adjusted probabilities of the primary outcome at various surgeon annual volume thresholds

Surgeon annual volume category	Number of patients	Number of surgeons	Adjusted probability of outcome (95% CI)	Surgeon annual volume threshold	Population attributable risk (%)	Population impact number
0–4	593	306	63.8% (59.9 to 67.7)	-	-	-
5–9	425	54	45.8% (41.2 to 50.5)	$\geq$ 5	1.7%	60.5
10–14	831	56	42.4% (39.0 to 45.7)	$\geq$ 10	1.9%	53.0
15–19	1172	46	44.5% (41.8 to 47.3)	$\geq$ 15	2.2%	44.9
20–24	1737	47	41.4% (39.1 to 43.7)	$\geq$ 20	2.9%	34.6
25–29	1815	47	40.1% (37.9 to 42.3)	$\geq$ 25	3.7%	26.8
30–34	1516	33	38.2% (35.9 to 40.6)	$\geq$ 30	4.7%	21.5
35–39	1626	26	35.1% (32.8 to 37.3)	$\geq$ 35	5.4%	18.5
40–44	1449	20	36.7% (34.3 to 39.1)	$\geq$ 40	6.0%	16.6
45–49	1237	14	32.6% (30.1 to 35.1)	$\geq$ 45	6.9%	14.5
50–59	2637	26	29.8% (28.0 to 31.5)	$\geq$ 50	7.4%	13.6
60–69	1345	12	38.3% (35.8 to 40.9)	$\geq$ 60	7.9%	12.7
70–79	1040	8	26.2% (23.6 to 28.9)	$\geq$ 70	10.2%	9.8
80–89	1710	10	32.1% (29.9 to 34.4)	$\geq$ 80	10.3%	9.7
90–99	1213	6	20.8% (18.4 to 23.1)	$\geq$ 90	13.8%	7.3
100 and over	2040	7	22.7% (21.0 to 24.5)	$\geq$ 100	13.1%	7.6

Note: Based on analysis of 22,406 unique patients with a surgeon listed

thyroid surgeon”, although they did not specify how high volume should be defined [11]. Earlier guidance from the American Thyroid Association [12] cited the work of Kandil et al. [19] (using HCUP-NIS) and their definition of high-volume surgery at  $>$  100 annual thyroidectomies. The authors recognised that the relatively low number of surgeons performing this volume (18.5% of surgeons in their sample) and the concentration of these surgeons in large

urban centres made this an unrealistic target [12]. Results from the present study, although not definitive, provided some evidence of a plateau in the improvement in outcomes when surgical volumes exceed 90 per annum.

A recent positional statement by the European Society of Endocrine Surgeons also concluded that a surgeon volume-outcome effect exists for thyroid surgery [20]. They identify  $>$  50 procedures per annum per surgeon as an

appropriate threshold for high volume and < 25 procedures per annum per surgeon as an appropriate threshold for low volume. However, the authors note the lack of high-quality evidence within the research literature, the dominance of studies from the USA and the potential for the findings from observational studies to be biased by unrepresentative patient cohorts, local healthcare funding and organisational models and selective referral.

A volume-outcome relationship can arise where larger volumes lead to more skilled and more efficient practice. In a setting where healthcare provision is based on health insurance schemes, selective referral can make it more likely that a volume-outcome relationship will be observed [21, 22]. Centres achieving better outcomes may receive more referrals, a guarantee of larger volumes [23]. The English public does not have great flexibility of choice of their healthcare provider [24]. As such, selective referral is unlikely to drive any observed volume-outcome relationship in our study. As our study includes all patients seen by the NHS in England, our cohorts are representative of the situation in England, although extrapolation to other settings should be undertaken with caution.

The link between volume and secondary outcomes provides some insight as to specific complications arising from thyroidectomy. At 1-year follow-up, the presence of hypoparathyroidism, vocal cord palsy, and stridor and the requirement for tracheostomy were significantly associated with a lower surgeon annual volume. That we identified a relationship for all outcomes investigated is striking compared to previous research [5–7, 9, 19]. All these complications are life-changing, and our findings emphasise the importance of the volume-outcome relationship at a patient level. Hypoparathyroidism, vocal cord palsy, and tracheostomy occurred in 11.3%, 2.4%, and 1.4% of patients respectively at 1-year follow-up. It is likely that vocal cord palsy is underreported in HES, being either not recognised or not reported in the immediate post-operative period. Stridor however will be both diagnosed and documented at an early stage due to patient breathing difficulty.

Permanent hypoparathyroidism, vocal cord palsy, stridor, and tracheostomy were more common following total thyroidectomy than lobectomy which supports the observation by Hauch et al. [13]. Our decision to focus on outcomes for total thyroidectomy provided a relatively homogenous dataset whilst also focussing on the procedure with the greatest potential patient impact.

A consistent relationship between hospital volume and outcomes was not identified. Although the impact of institutional volume on outcomes of thyroid surgery has been less well studied, this supports previous research [19]. This suggests that it is the judgement and skill of the operating surgeon rather than the organisation of the wider surgical team that contributes most to patient outcomes for total

thyroidectomy in England. There has been some debate as to which factors are responsible for the hospital volume-outcome effect (e.g. organisation structure, staff training, equipment availability, and the directionality of the relationship (with selective referral leading to larger volume)) [22, 25]. For surgeon volume, the reason for better outcomes in larger volume surgeons are, perhaps, more clearly related to experience and evidence of practice making perfect [25].

In comparison with data from the UKRETS audit, average annual recorded volumes for total thyroidectomy were much smaller in HES. However, the number of total thyroidectomies recorded in UKRETS was less than 60% of that recorded in HES and the number of surgeons recording total thyroidectomies in UKRETS was only one-third of the number of surgeons identified as conducting total thyroidectomies in HES. Thus, it is reasonable to conclude that UKRETS contains a disproportionately large number of high-volume surgeons relative to HES and that average outcomes reported in UKRETS will be better than those described here. Reporting of thyroid surgery to UKRETS is now required of surgeons performing thyroid procedures in England. Although our data suggest that many surgeons are not compliant with this requirement, if enforced, it could be a key driver to reducing low-volume thyroid surgery in England.

## Strengths

In addition to the strengths already discussed, our study uses recent data from all thyroid surgery in England. The vast majority of thyroid surgery in England is funded by the NHS and the HES database represents one of the most complete healthcare datasets globally. To maximise the homogeneity of this dataset, total thyroidectomy was used as an exemplar condition. Total thyroidectomy has a higher complication rate than lobectomy, so more suited to the study of volume-outcome relationships [13].

## Limitations

The HES dataset was unable to adjust for aspects of clinical presentation, such as the stage of cancer at the point of the procedure, and it is possible that presentation could vary by volume. However, it is difficult to assess in which direction any bias may be and whether it would be systematic. It could be argued that larger volume surgeons may see more complicated cases, but conversely that patients with multiple morbidity may be more likely to be operated on by surgeons from different specialities, conducting low volumes of thyroidectomies.

By removing laryngectomies and head and neck cancer patients from our dataset, we have tried to minimise these impacts. We did not adjust for the presence of pre-surgery

vocal cord palsy, stridor, or hypocalcaemia; however, they are likely to be rare. HES also contains only limited information on the procedure undertaken, such as the use of intra-operative neuromonitoring. Coding errors within HES are known to exist [26]. For this study, this is likely to result in an underestimation of complication rates. Non-clinical coders, who enter data into HES do not identify hypocalcaemia unless it is specifically written into the medical notes. Furthermore, stridor, vocal cord palsy, hypocalcaemia, and hypoparathyroidism would only be recorded if they resulted in a hospital admission. Thus, they are likely to represent a substantial underreport of the true rate of occurrence of these complications, as experienced by patients [27]. Nevertheless, they will represent the incidence of the most severe forms of these conditions. For this reason, we chose a surrogate measure for complications as our primary outcome measure. We considered that this would more accurately reflect the occurrence of a complication following total thyroidectomy.

Although both length of stay and early emergency readmission may be a function of post-operative management processes, poor post-operative care planning, leading to early readmission or extended stay, is also a valid marker for poor outcomes for both the patient and the health service. Updated coding guidance for HES and improved compliance with the UKRETS clinical audit will help improve data quality.

## Conclusions

Our study supports the view that consolidation of total thyroidectomy surgery within larger volume surgeons would improve patient outcomes in England. A clear volume threshold has not been identified but reducing low-volume surgery should be a priority. There was limited evidence of an association between hospital volume and outcomes, consistent with previous research.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00423-021-02223-8>.

**Acknowledgements** We acknowledge NHS Digital for permission to use their data in this report. The GIRFT programme is providing a framework for examining contemporary clinical practice in unprecedented detail and breadth. We also thank all staff within individual NHS trusts who collected and entered the data used in this study. The study protocol was not pre-registered.

**Authors' contributions** This study was designed and organised by ML, SA, NA, JD, and WKG. Data cleaning, analysis, and writing of the first draft were by WKG, supported by SA, NT, ML, and JD. All authors critically reviewed the manuscript and agreed to submission of the final draft. Study conception and design—ML, SA, NA, JD, and WKG; acquisition of data—WKG, supported by SA and JD; analysis and interpretation of data—WKG, supported by SA, NT, ML, and JD;

drafting of manuscript—WKG, SA, NT, and JD; critical revision of manuscript—WKG, ML, NT, SA, and JD.

**Availability of data** This report does not contain patient identifiable data. Data in this report are anonymised. Request for any underlying data will not be granted as the data are calculated from data under licence/data sharing agreement from NHS Digital and/or other data provider where conditions of use (and further use) apply. Copyright© 2013, 2014, 2015, 2016, 2017 re-used with the permission of the Health & Social Care Information Centre. All rights reserved.

## Declarations

**Ethics approval** Ethical approval was not sought for the present study because it did not directly involve human participants. This study was completed in accordance with the Helsinki Declaration as revised in 2013.

**Informed consent to participate** Informed consent was not sought for the present study because it was an analysis of routine clinical data.

**Consent for publication** All authors consent to publication.

**Conflict of interest** The authors declare no competing interests.

## References

1. Bhattarai N, McMeekin P, Price C et al (2016) Economic evaluations on centralisation of specialised healthcare services: a systematic review of methods. *BMJ Open*. 6:e011214
2. Imison C, Sonola L, Honeyman M et al (2014) The reconfiguration of clinical services. What is the evidence? The King's Fund, London
3. 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* 94:145–161
4. Nicholl J, West J, Goodacre S et al (2007) The relationship between distance to hospital and patient mortality in emergencies: an observational study. *Emerg Med J* 24:665–668
5. Nouraei SA, Virk JS, Middleton SE et al (2017) A national analysis of trends, outcomes and volume-outcome relationships in thyroid surgery. *Clin Otolaryngol* 42:354–365
6. Aspinall S, Oweis D, Chadwick D (2019) 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). *Langenbecks Arch Surg* 404:421–430
7. Adkisson CD, Howell GM, McCoy KL et al (2014) Surgeon volume and adequacy of thyroidectomy for differentiated thyroid cancer. *Surgery*. 156:1453–59 (discussion 60)
8. Adam MA, Thomas S, Youngwirth L et al (2017) Is there a minimum number of thyroidectomies a surgeon should perform to optimize patient outcomes? *Ann Surg* 265:402–407
9. Meltzer C, Hull M, Sundang A et al (2019) Association between annual surgeon total thyroidectomy volume and transient and permanent complications. *JAMA Otolaryngol Head Neck Surg* 145:830–837
10. Liang TJ, Liu SI, Mok KT et al (2016) Associations of volume and thyroidectomy outcomes: a nationwide study with

- systematic review and meta-analysis. *Otolaryngol Head Neck Surg* 155:65–75
11. Patel KN, Yip L, Lubitz CC et al (2020) The American Association of Endocrine Surgeons guidelines for the definitive surgical management of thyroid disease in adults. *Ann Surg* 271:e21–e93
  12. Haugen BR, Alexander EK, Bible KC et al (2016) 2015 American Thyroid Association management guidelines for adult patients with thyroid nodules and differentiated thyroid cancer: the American Thyroid Association guidelines task force on thyroid nodules and differentiated thyroid cancer. *Thyroid* 26:1–133
  13. Hauch A, Al-Qurayshi Z, Randolph G et al (2014) Total thyroidectomy is associated with increased risk of complications for low- and high-volume surgeons. *Ann Surg Oncol* 21:3844–3852
  14. NHS Digital (2018) Hospital Episode Statistics (HES) analysis guide. NHS Digital, London
  15. Information Standard Board for Health and Social Care (2013) Anonymisation standard for publishing health and social care data specification (Process Standard). NHS Digital, London
  16. Gilbert T, Neuburger J, Kraindler J et al (2018) Development and validation of a Hospital Frailty Risk Score focusing on older people in acute care settings using electronic hospital records: an observational study. *Lancet* 391:1775–1782
  17. Nimptsch U, Mansky T (2017) Hospital volume and mortality for 25 types of inpatient treatment in German hospitals: observational study using complete national data from 2009 to 2014. *BMJ Open*. 7:e016184
  18. British Association of Endocrine and Thyroid Surgeons (2017) Fifth National Audit Report. Dendrite Clinical Systems Ltd, Henley-on-Thames
  19. Kandil E, Noureldine SI, Abbas A et al (2013) The impact of surgical volume on patient outcomes following thyroid surgery. *Surgery*. 154:1346–52 (discussion 52-3)
  20. Lorenz K, Raffaelli M, Barczynski M et al (2020) Volume, outcomes, and quality standards in thyroid surgery: an evidence-based analysis-European Society of Endocrine Surgeons (ESES) positional statement. *Langenbecks Arch Surg* 405:401–425
  21. Birkmeyer JD, Finlayson EV, Birkmeyer CM (2001) Volume standards for high-risk surgical procedures: potential benefits of the Leapfrog initiative. *Surgery* 130:415–422
  22. Luft HS, Hunt SS, Maerki SC (1987) The volume-outcome relationship: practice-makes-perfect or selective-referral patterns? *Health Serv Res* 22:157–182
  23. Halm EA, Lee C, Chassin MR (2002) Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med* 137:511–520
  24. Dixon A, Robertson R, Appleby J et al (2010) Patient choice: how patients choose and how providers respond. The King's Fund, London
  25. Hentschker C, Mennicken R (2018) The volume-outcome relationship revisited: practice indeed makes perfect. *Health Serv Res* 53:15–34
  26. Nouraei SA, Hudovsky A, Frampton AE et al (2015) A study of clinical coding accuracy in surgery: implications for the use of administrative big data for outcomes management. *Ann Surg* 261:1096–1107
  27. Abdelhamid A, Aspinall S (2021) Intraoperative nerve monitoring in thyroid surgery: analysis of United Kingdom registry of endocrine and thyroid surgery database. *Br J Surg* 108:182–187

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.