Systematic Review and Meta-analysis of Robotic vs Conventional Thyroidectomy Approaches for Thyroid Disease

Gordon H. Sun, Lilia Peress and Melissa A. Pynnonen

Otolaryngology -- Head and Neck Surgery 2014 150: 520 originally published online 5 February 2014
DOI: 10.1177/0194599814521779

The online version of this article can be found at:
http://oto.sagepub.com/content/150/4/520

Additional services and information for Otolaryngology -- Head and Neck Surgery can be found at:

Email Alerts: http://oto.sagepub.com/cgi/alerts
Subscriptions: http://oto.sagepub.com/subscriptions
Reprints: http://www.sagepub.com/journalsReprints.nav
Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Apr 1, 2014

OnlineFirst Version of Record - Feb 5, 2014
What is This?
Systematic Review and Meta-analysis of Robotic vs Conventional Thyroidectomy Approaches for Thyroid Disease

Gordon H. Sun, MS, MD1,2,3*, Lilia Peress1*, and Melissa A. Pynnonen, MSc, MD1

Abstract

Objective. This study compared postoperative technical, quality-of-life, and cost outcomes following either robotic or open thyroidectomy for thyroid nodules and cancer.

Data Sources. PubMed, Ovid MEDLINE, EMBASE, ISI Web of Science, and the Cochrane Central Register of Controlled Trials.

Review Methods. We examined relevant controlled trials, comparative effectiveness studies, and cohort studies for eligible publications. We calculated the pooled relative risk for key postoperative complications, mean differences for operative time, and standardized mean differences for length of stay (LOS) using random effects models. Quality-of-life outcomes were summarized in narrative form.

Results. The meta-analysis comprised 11 studies with 726 patients undergoing robotic transaxillary or axillo-breast thyroidectomy and 1205 undergoing open thyroidectomy. There were no eligible cost-related studies. Mean operative time for robotic thyroidectomy exceeded open thyroidectomy by 76.7 minutes, while no significant difference in LOS was identified. There were no significant differences in hematoma, seroma, recurrent laryngeal nerve injury, hypocalcemia, or chyle leak rates. The systematic review included 12 studies. Voice, swallowing, pain, and paresthesia outcomes showed no significant differences between the 2 approaches. The robotic cohort reported higher cosmetic satisfaction scores, although follow-up periods did not exceed 3 months and no validated questionnaires were used.

Conclusions. Transaxillary and axillo-breast robotic and open thyroidectomy demonstrate similar complication rates, but robotic approaches may introduce the risk of new complications and require longer operative times. Robotic thyroidectomy appears to improve cosmetic outcomes, although longer follow-up periods and use of validated instruments are needed to more rigorously examine this effect.

Keywords

robotic surgery, thyroid nodule, thyroid cancer, length of stay, operative time, quality of life, hemorrhage, recurrent laryngeal nerve injury, hoarseness, hypoparathyroidism, hypocalcemia, brachial plexus injury

Received August 16, 2013; revised January 6, 2014; accepted January 9, 2014.

The traditional surgical approach for thyroidectomy, the transcervical incision popularized by Kocher, remained relatively unchanged from its inception in the late 19th century until 1997, when an endoscopic approach was described.1-3 Subsequently, in 2000, the da Vinci surgical robot (Intuitive Surgical, Inc, Sunnyvale, California) was approved by the US Food and Drug Administration (FDA) for certain laparoscopic surgical procedures.4 The surgical robot allows the surgeon to perform the operation by manipulating the robot arms while seated at a control panel, without the need for the surgeon to hold and manipulate an endoscopic or other surgical instrument by hand, thus reducing tremor and improving fine motor control of instrumentation. Since then, the surgical robot has been used to develop minimally invasive approaches to thyroidectomy, with or without complementary endoscopy. However, the use of the surgical robot for thyroidectomy is not FDA approved and remains off-label.

There is substantial controversy over the role of the robot in thyroidectomy because traditional open thyroidectomy already has a low morbidity rate and produces excellent results. Improved cosmesis is the primary advantage of robotic thyroidectomy, which trades a cervical incision for one in the axillary space or anterior chest wall.5,6 Other
advantages may include improved surgeon ergonomics and dexterity. Limitations of robotic thyroidectomy include longer operative time, increased equipment and staffing needs, and complications related to the surgical approach, such as brachial plexus injury.

A previous meta-analysis compared robotic, endoscopic, and conventional open thyroidectomy approaches in 9 studies, concluding that the robotic approach was safe and efficacious. However, the study examined only pain and cosmesis and did not study other important quality-of-life (QoL) outcomes such as voice and swallowing function. Another recent meta-analysis of 6 studies compared robotic and endoscopic thyroidectomies and concluded that the risk of complications was higher using the robotic approach, but the authors did not define the nature of the complications of interest.

We conducted the following meta-analysis to examine postoperative complications and provide a more comprehensive overview of QoL-related issues in clinical trials and cohort studies comparing robotic and conventional open thyroidectomy for thyroid cancer and nodules. Our secondary objective was to characterize the methodological quality of the literature.

**Patients and Methods**

**Search Strategy**

The findings from this systematic review and meta-analysis were reported in accordance with the 2009 Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines. The protocol was registered with the PROSPERO database (#CRD42012003331). On January 2, 2013, we searched PubMed, Ovid MEDLINE, EMBASE, ISI Web of Science, and the Cochrane Central Register of Controlled Trials. We also manually searched the bibliographies of published review articles and other relevant publications for additional studies. Due to limited resources, we did not review conference proceedings and abstracts, grey literature databases, and non-English publications. The full search strategy is available in the Appendix (available at otojournal.org) and at PROSPERO (http://www.crd.york.ac.uk/PROSPERO/display_record.asp?ID=CRD42012003331). The complete search yielded 670 references. After duplicate entries were removed, 277 references remained. The references were imported into EndNote (Thomson Reuters, New York, New York) for abstract review. This study was exempt from review by the University of Michigan Medical School Institutional Review Board.

**Study Selection**

Two authors (G.H.S. and L.P.) screened the titles and abstracts of the 277 articles, with disagreements being resolved by consensus or arbitration with a third author (M.A.P.). We included randomized and nonrandomized controlled trials, comparative effectiveness studies, and prospective or retrospective cohort studies in which robotic and conventional open approaches to the thyroid were compared in patients with thyroid nodules or cancer. We included studies using endoscopic thyroidectomy techniques only if the endoscope was used in the surgical robot arm and the comparison group was open thyroidectomy. According to our a priori protocol, all eligible studies were required to report any one of the following: (1) at least 1 commonly accepted postthyroidectomy health-related outcome such as surgical complications, operative time, or hospital length of stay (LOS); (2) at least 1 QoL measure such as voice, swallowing, or cosmetic satisfaction; or (3) actual treatment costs for robotic and open thyroidectomy patients. We excluded studies for which a full-length manuscript was not available (eg, conference abstracts).

**Data Extraction and Publication Quality Assessment**

The same 2 authors independently abstracted the following variables: number of patients; patient age, sex, race/ethnicity, and country where the study took place; body mass index (BMI); benign vs malignant pathology; tumor size or extent of disease; and total vs partial thyroidectomy, neck dissection, and surgical approach. We recorded complications, robot-to-open conversion rate, operative time, LOS, and treatment costs. We evaluated publication quality according to the Martin criteria, which are 10 elements believed to be essential to allow comparison of surgical outcomes across practices and institutions.

**Systematic Review Approach and Statistical Analysis**

We described patient demographic and tumor characteristics using descriptive statistics. We compared complication rates between groups using 2-sample *t* tests and tests of proportions. For meta-analysis, we calculated pooled relative risk (RR) with 95% confidence intervals (CIs) to summarize categorical complications. We compared operative times using mean differences with 95% CIs and hospital LOS with standardized mean differences (SMDs). The SMD is defined as the difference between the mean outcomes of the compared groups divided by the pooled standard deviation of the outcome. We used DerSimonian and Laird random effects models to pool outcomes across studies. We used forest plots to depict quantitative outcomes and Cochran’s *Q* and *I*² tests to examine heterogeneity between studies. We conducted all statistical analyses with Stata 12.1 (StataCorp LP, College Station, Texas) and used 2-sided (*P* < .05) tests of significance.

**Results**

**Study Characteristics**

We reviewed the titles and abstracts and excluded 261 of the 277 articles (Figure 1). We reviewed full-length manuscripts of the remaining 16 publications and excluded 4 more studies. We excluded Broome et al because the costs for robotic surgery were not based on the authors’ original data but instead extrapolated from 2 previously published studies. We excluded Cabot et al because the study reported outcomes selectively and Lee et al because no patient outcomes were reported. We excluded Yoo et al because the diagnoses were unclear (eg, thyroid nodule vs thyroiditis). Twelve publications were included in the
systematic review (Table 1 and online Appendix). None of the articles described cost.

We contacted study authors to obtain or clarify data. From Dr Kyung Tae (Hanyang University, Seoul, South Korea), we obtained previously unpublished individual-level data spanning October 2008 to October 2010 (personal communication, May 25, 2013), which encompassed 2 published studies from his group that had overlapping cohorts. Altogether, there were 97 robotic and 307 open cases from Tae’s group. We used these unpublished data rather than the individual published studies from his group to avoid case duplication. Another article from Tae’s group was included in the systematic review but not the meta-analysis because it included only outcomes related to QoL.

The 11 articles in the meta-analysis reported characteristics and outcomes for 1931 patients: 726 patients undergoing robotic thyroidectomy and 1205 patients undergoing open thyroidectomy. Table 2 compares the demographic and surgical characteristics of all patients by surgical approach. There were significant differences in age, sex, BMI, and the proportion undergoing total thyroidectomy. The robotic group was younger (40.5 vs 49.2 years), had more women (91.2% vs 78.5%), had a lower mean BMI (23.1 vs 24.2), and underwent fewer total thyroidectomies (58.1% vs 75.1%).

Methodological Quality and Assessment of Bias

Using Martin criteria, we analyzed 8 articles for publication quality (0-10, with higher scores indicating higher quality), finding that the mean score of the studies was 7.75 (median, 8; range, 3-9). All studies described the method of patient accrual, morbidity rate, and complications; however, definitions of complications varied considerably. For example, Lee et al defined permanent vocal cord paralysis as no evidence of recovery within 6 months, while Lee et al used 3 months as the cutoff. Follow-up duration was reported in 5 of 8 studies. The longest period of follow-up was 29.1 months (range, 10-38 months), while the other studies followed patients from 3 months up to 1 year.

Patient Outcomes

Operative time. Seven articles reported overall operative time, while 4 articles reported time separately for subtotal vs total thyroidectomies. Operative time was significantly longer in the robotic thyroidectomy group (Figures 2-4). The mean operative time of robotic thyroidectomy exceeded conventional open thyroidectomy by 76.7 minutes (95% confidence interval [CI], 47.3-106.1). There was significant heterogeneity among the studies (I² = 95.3%, P < .001). There was an increased mean difference of 48.1 minutes (95% CI, 33.6-62.7) via robotic approaches specifically in total thyroidectomy patients and 37.3 minutes (95% CI, 12.0-62.5) via robotic approaches in subtotal thyroidectomy patients. There was significant heterogeneity noted in both of these comparisons (I² = 84.8%, P = .001 for total thyroidectomy; I² = 97.3%, P < .001 for subtotal thyroidectomy).
Seven articles 16,17,20-24 reported LOS. The SMD between open and robotic thyroidectomy hospital LOS was 2.006 (95% CI, 2.25 to 0.24), meaning that robotic thyroidectomy patients had a small, nonsignificant lower mean LOS compared with their open thyroidectomy counterparts (Figure 5). Significant heterogeneity among studies was observed ($I^2 = 75.1\%$, $P = .001$).

### Table 1. Characteristics of All 12 Studies Incorporated into Systematic Review.

<table>
<thead>
<tr>
<th>Study (Author, Year)</th>
<th>Study Type</th>
<th>Country</th>
<th>Total No. of Patients (Open, Robotic)</th>
<th>Robotic Approach</th>
<th>Follow-up Period</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foley et al,19 2012</td>
<td>Retrospective</td>
<td>United States</td>
<td>27 (16, 11)</td>
<td>Unilateral transaxillary, with endoscope, gasless</td>
<td>Perioperative</td>
</tr>
<tr>
<td>Kang et al,20 2012</td>
<td>Retrospective</td>
<td>South Korea</td>
<td>165 (109, 56)</td>
<td>Unilateral transaxillary, with endoscope, gasless</td>
<td>Tg at 6 mo, US at 1 y</td>
</tr>
<tr>
<td>Kim et al,26 2013</td>
<td>Prospective</td>
<td>South Korea</td>
<td>37 (18, 19)</td>
<td>Bilateral axillo-breast, with endoscope and insufflation</td>
<td>Perioperative</td>
</tr>
<tr>
<td>Kim et al,21 2011</td>
<td>Retrospective</td>
<td>South Korea</td>
<td>207 (138, 69)</td>
<td>Bilateral axillo-breast, with endoscope and insufflation</td>
<td>Perioperative</td>
</tr>
<tr>
<td>Landry et al,22 2012</td>
<td>Retrospective</td>
<td>United States</td>
<td>50 (25, 25)</td>
<td>Unilateral transaxillary, with endoscope; use of insufflation not addressed</td>
<td>Perioperative</td>
</tr>
<tr>
<td>Lee et al,25 2012</td>
<td>Prospective</td>
<td>South Korea</td>
<td>88 (46, 42)</td>
<td>Unilateral transaxillary, without endoscope, gasless</td>
<td>1 wk, 3 mo</td>
</tr>
<tr>
<td>Lee et al,23 2010</td>
<td>Prospective</td>
<td>South Korea</td>
<td>84 (43, 41)</td>
<td>Unilateral transaxillary, with endoscope, gasless</td>
<td>1 wk, 3 mo</td>
</tr>
<tr>
<td>Lee et al, 201127</td>
<td>Prospective</td>
<td>South Korea</td>
<td>411 (237, 174)</td>
<td>Bilateral axillo-breast, with endoscope and insufflation</td>
<td>Tg at 8-12 wk</td>
</tr>
<tr>
<td>Lee et al,24 2012</td>
<td>Retrospective</td>
<td>South Korea</td>
<td>458 (266, 192)</td>
<td>Unilateral axillo-breast, with endoscope, gasless</td>
<td>Mean 29.1 mo (range, 10-38 mo)</td>
</tr>
<tr>
<td>Tae et al,16 2012a</td>
<td>Retrospective</td>
<td>South Korea</td>
<td>301 (226, 75)</td>
<td>Unilateral axillo-breast or axillary, with endoscope, gasless</td>
<td>Mean 11.18 mo in robotic, 12.53 mo in open</td>
</tr>
<tr>
<td>Tae et al,18 2012b</td>
<td>Prospective</td>
<td>South Korea</td>
<td>111 (61, 50)</td>
<td>Unilateral axillo-breast, with endoscope, gasless</td>
<td>1 d, 1 wk, 1 mo, 3 mo, 6 mo</td>
</tr>
<tr>
<td>Tae et al,17 2011a</td>
<td>Retrospective</td>
<td>South Korea</td>
<td>208 (167, 41)c</td>
<td>Unilateral axillo-breast (2/41 axillary without breast port), with endoscope, gasless</td>
<td>1 wk, 1 mo, 3 mo</td>
</tr>
</tbody>
</table>

**Abbreviations:** Tg, thyroglobulin; US, ultrasound.

*These studies, plus additional information obtained directly from the corresponding author, were combined for meta-analysis.

*This was excluded from meta-analysis due to reporting of only quality-of-life outcomes.

*There were 167 open cases in the Methods section but only 163 in Table 1 of the Results section of the referenced article.

### Table 2. Demographic and Surgical Characteristics of Patients from All Eligible Studies.a

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Robotic Cohort (n = 726)</th>
<th>Open Cohort (n = 1205)</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean age, y</td>
<td>40.5</td>
<td>49.2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Female sex, No. (%)</td>
<td>662 (91.2)</td>
<td>947 (78.5)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Body mass indexb</td>
<td>23.1</td>
<td>24.2</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Thyroid malignancy, No. (%)c</td>
<td>660/707 (93.4)</td>
<td>1111/1187 (93.6)</td>
<td>.83</td>
</tr>
<tr>
<td>Total thyroidectomy, No. (%)</td>
<td>422 (58.1)</td>
<td>905 (75.1)</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Tumor size, mmd</td>
<td>7.9</td>
<td>8.0</td>
<td>.64</td>
</tr>
</tbody>
</table>

aNot all patient characteristics were reported in every study. $P < .05$ represents the threshold of statistical significance.

bBased on 298 robotic and 434 open cases.

cBased on 707 robotic and 1187 open cases.

Hospital length of stay. Seven articles 16,17,20-24 reported LOS. The SMD between open and robotic thyroidectomy hospital LOS was $-0.006$ (95% CI, $-0.25$ to $0.24$), meaning that robotic thyroidectomy patients had a small, nonsignificant lower mean LOS compared with their open thyroidectomy counterparts (Figure 5). Significant heterogeneity among studies was observed ($I^2 = 75.1\%$, $P = .001$).
Postoperative complications. There were no statistically significant differences between robotic and open thyroidectomy outcomes in the rates of hematoma, seroma, recurrent laryngeal nerve (RLN) injury, hypocalcemia, or chyle leak (Figures 6-12). There was no significant heterogeneity among studies for any complication except temporary hypocalcemia ($I^2 = 83.7\%, P < .001$). There were no deaths or tracheal injuries reported in any of the studies.

Conversion from robotic to open thyroidectomy was necessary in 2 of 243 (0.8\%) cases. The 6 articles that evaluated brachial plexus injury reported a combined event rate of 8 of 366 (2.2\%). Foley et al reported 1 patient with discomfort on shoulder movement that resolved prior to the first postoperative visit. Landry et al reported 1 patient with a decreased range of motion due to pain and a second patient with trouble gripping...
and handwriting due to brachial plexus injury; both patients recovered within several months. Lee et al\textsuperscript{23} reported brachial plexus injuries in 5 of 41 (12.2\%) patients, as manifested by shoulder discomfort, 1 week postoperatively; symptoms resolved in all patients by 3 months.

### Quality of Life

#### Voice

Lee et al\textsuperscript{23} used the Voice Handicap Index\textsuperscript{28} preoperatively and 1 week and 3 months postoperatively for patient self-assessment of voice impairment. No statistically significant difference in voice symptom scores was found at any time point between the 2 groups.

In a separate study, Lee et al\textsuperscript{25} used both subjective and objective methods to evaluate patients preoperatively and 1 week and 3 months postoperatively. Subjective evaluation, performed by an experienced voice therapist using the GRBAS (Grade, Roughness, Breathiness, Asthenia, Strain) scale,\textsuperscript{29} found no difference at any of the 3 time points. Objective evaluation was performed using videolaryngostroboscopy

---

**Figure 4.** Subtotal thyroidectomy operative time, robotic vs open approach. CI, confidence interval; WMD, weighted mean difference.

**Figure 5.** Hospital length of stay, robotic vs open approach. CI, confidence interval; SMD, standardized mean difference.
(VLS), acoustic voice analysis, the Voice Range Profile (maximum and minimum frequencies, frequency range, intensity range, and maximum phonation time), and electroglottography (EGG). There were very few between-group differences identified, and all were temporary. In the acoustic voice analysis, the only significant difference was increased shimmer in the robotic group at 1 week, which resolved by 3 months. In addition, the aerodynamic studies and EGG revealed no notable differences at any time.

Tae et al\textsuperscript{18} used a 5-item questionnaire to investigate symptoms of vocal fatigue, hoarseness, pitch limitation, breathiness or weakness, and difficulty singing, with patients rating the severity of each symptom on a 5-point scale (none, mild, moderate, severe, or very severe). Preoperatively, the open cohort and the robotic cohort were similar in their reporting of vocal impairment. Postoperatively, the open cohort had a higher severity of voice symptoms at postoperative day 1 ($P = .008$), month 1 ($P = .049$), and month 3 ($P = .043$) but not at month 6.
Objective studies included VLS, acoustic analysis, and the Voice Range Profile. The VLS results were not reported. Acoustic analysis revealed no significant differences between the 2 groups at any time point. Voice Range Profile parameters found a temporary broader frequency range and a higher maximum frequency at 3 months in the robotic cohort. This difference resolved by 6 months postoperatively. All 111 patients completed the questionnaire and underwent acoustic voice analysis preoperatively and at 1 week postoperatively. However, due to loss to follow-up, only 105 patients completed the questionnaire at 1 month, 97 at 3 months, and 92 at 6 months, while 102 patients completed voice analysis at 1 month, 92 at 3 months, and 85 at 6 months postoperatively.

Swallowing. Swallowing was described in 2 articles with discordant results. Lee et al use the validated 6-item Swallowing Impairment Index (SIS-6) for the assessment of increased effort of swallowing, choking, and throat clearing. Patients in both the robotic and open cohorts had similar
Preoperatively, the open cohort had significantly worse mean SIS-6 scores at both 1 week ($P = .001$) and 3 months ($P = .007$).

Tae et al$^{18}$ created a 3-item questionnaire to evaluate pain or difficulty swallowing, foreign body sensation, and choking or coughing when swallowing. Patients reported their severity of symptoms on a 5-point scale (no symptoms, mild, moderate, severe, or very severe). There were no between-group differences in swallowing impairment at any time point measured up to 6 months.

Cosmetic satisfaction. Three articles described cosmetic satisfaction using nonvalidated instruments. Lee et al$^{23}$ assessed cosmetic satisfaction at 3 months postoperatively using a 5-point scale (extremely satisfied, satisfied, acceptable, dissatisfied, or extremely dissatisfied) and found that patients in the robotic group reported significantly greater satisfaction than those in the open group ($P < .0001$). Twenty-four (58.5%) patients in the robotic group reported being extremely satisfied compared with 5 (11.6%) patients in the open group. No patients in the robotic group reported being...
either extremely dissatisfied or dissatisfied, while 8 patients (18.6%) in the open group reported being dissatisfied and 1 (2.3%) reported being extremely dissatisfied.

Tae et al\textsuperscript{17} assessed cosmetic satisfaction using a 5-point scale (very satisfied to very dissatisfied) at 1 week, 1 month, and 3 months postoperatively. Mean satisfaction was significantly better in the robotic group ($P < .001$ for each time point). However, the 2 study arms suffered from varying degrees of loss to follow-up; 32 of 41 (78.0%) patients in the robotic group reported cosmetic satisfaction, compared with only 89 of 163 (54.6%) in the open thyroidectomy group.

In 2012, Tae et al\textsuperscript{16} measured cosmetic satisfaction using the same 5-point scale at 1 day, 1 week, 1 month, and 3 months postoperatively. Mean satisfaction was significantly better in the robotic group than in the open group ($P < .001$ for each time point), but this study also suffered from differential loss to follow-up between the 2 study arms. In the robotic group, 48 of 75 (64.0%) patients reported cosmetic satisfaction scores, compared with 71 of 226 (31.4%) in the open thyroidectomy group.

Pain. Three articles described postoperative pain. Lee et al\textsuperscript{23} used a 5-point grading scale to grade neck and anterior chest pain at 24 hours postoperatively. Pain was then reported as none, very slight, slight, moderate, or severe. Moderate or severe pain was reported by 5 (12.2%) patients in the robotic group and 6 (14.0%) in the open group ($P = .43$). The authors used an identical postoperative analgesic regimen for the 2 groups and found no difference in the number of postoperative analgesics used in each cohort ($P = .76$).

In the 2011 study by Tae et al\textsuperscript{17}, 121 patients scored their postoperative neck and anterior chest pain on a 5-point scale (none, mild, moderate, severe, or very severe) at 1 week, 1 month, and 3 months postoperatively. There was no significant difference in the severity of neck pain. However, the robotic cohort reported significantly higher anterior chest pain scores at 1 week ($P < .001$). This difference resolved by 1 month postoperatively. In 2012, the same authors\textsuperscript{16} reported 119 patients who scored postoperative neck and anterior chest pain using the same 5-point scale but at 4 time points: 1 day, 1 week, 1 month, and 3 months postoperatively. There was again no significant difference in neck pain at any time point measured, but the robotic group had significantly higher anterior chest pain scores at 1 day, 1 week, and 1 month postoperatively ($P < .001$, $P < .001$, and $P = .014$, respectively). This difference in chest pain resolved by the 3-month postoperative visit.

Paresthesia. Three articles reported postoperative paresthesia in patients. Foley et al\textsuperscript{19} reported symptoms in 2 of 11 (18.2%) patients who underwent robotic thyroidectomy, both of whom recovered by the first postoperative visit. Lee et al\textsuperscript{23} assessed for the presence of paresthesia or hyperesthesia of the neck with a questionnaire completed by all participants at 1 week and 3 months postoperatively. Significantly more patients reported neck paresthesia in the open cohort than in the robotic cohort, both at 1 week (95.3% vs 36.6%, $P = .010$) and at 3 months (65.1% vs 9.8%, $P = .002$) postoperatively. Anterior chest paresthesia at 1 week was reported by 19 (46.3%) patients in the robotic group and 1 (2.4%) in the open group. At 3 months, 8 (19.5%) patients in the robotic group and none in the open group were still experiencing chest paresthesia.

In the 2012 study by Tae et al,\textsuperscript{16} 119 patients reported paresthesias using a 4-point scale at 1 day, 1 week, 1 month, and 3 months postoperatively. The severity of neck paresthesias was significantly different at 3 months only, with the open group reporting a greater degree of neck
paresthesias (P = .012). The severity of anterior chest paresthesias was significantly greater in the robotic group 1 day, 1 week, and 1 month after surgery (P = .001, P < .001, and P = .006, respectively). This difference resolved by 3 months postoperatively.

**Discussion**

The current meta-analysis supports many of the findings of the previous meta-analysis by Jackson et al. We found no difference in the rates of RLN injury, hypocalcemia, hema-

toma, seroma, and chyle leakage. We also found that robotic thyroidectomy cases had significantly longer operative times (76.7 minutes in our study vs 42.05 minutes in Jackson et al). To test whether this operative time discrepancy was due to robotic operative experience, we dropped the 2 US studies since South Korean surgeons generally have more experience with the surgical robot. This did not change the difference in operative times substantially (72.1 minutes longer in robotic thyroidectomy cases). Further analysis of the impact of the learning curve and institutional experience with the surgical robot on patient outcomes is warranted.

We caution that none of the studies in the meta-analysis were designed as a priori noninferiority or equivalence trials. Consequently, noninferiority of the experimental robotic technique cannot necessarily be assumed. This is supported in part by the fact that although many of the measured surgical outcomes were similar, the robotic approach has additional operative risks that have no analog in traditional open surgery. Such unique complications can also have additional downstream adverse effects on QoL, such as brachial plexus injury and its impact on the ability to perform manual labor. Future research examining the impact of this particular complication on overall QoL would be informative.

Other discordant results from our meta-analysis highlight the complexity of evaluating the robotic thyroidectomy technique. For example, our study did not detect a significant difference in the overall LOS between the robotic and open cohorts, in contrast to Jackson et al. Since the previous meta-analysis included only studies conducted in Asia and the average LOS for all causes in South Korea is approximately triple that of the United States, we conducted a sensitivity analysis excluding the US study by Landry et al. We determined that the SMD between open and robotic thyroidectomy hospital LOS was −0.053 (95% CI, −0.32 to 0.21), still a nonsignificant difference in mean LOS. This persistent difference might be due to the previous meta-analysis having double-counted overlapping cohorts by Tae et al, a problem we circumvented by acquiring the entire population of patients directly from the author and pooling them together into a single group.

One key issue is the impact of robotic thyroidectomy on cosmetic outcomes. Conceptually, robotic thyroidectomy might be expected to produce better cosmetic outcomes due to placement of the incision in a less visible area, and patients most likely to undergo thyroidectomy are younger women who may place a higher premium on cosmesis. Indeed, the cohort undergoing robotic thyroidectomy in this meta-analysis were nearly all female and almost 9 years younger on average than the cohort undergoing open thyroidectomy. Moreover, the studies we identified suggested that cosmetic satisfaction was significantly higher in the patients undergoing robotic thyroidectomy.

However, these cosmetic results should be considered carefully in light of several methodological concerns. None of the studies on cosmesis were randomized or used validated instruments such as the Patient Scar Assessment Questionnaire, which would standardize patient responses and allow more appropriate, less biased evaluation of the effectiveness of robotic vs open approaches. Since none of the studies were randomized, all are susceptible to self-selection bias; in other words, patients selecting a robotic treatment approach may be inherently biased toward greater cosmetic satisfaction because they selected that approach specifically to avoid a neck incision. Moreover, the results of the studies by Tae et al were confounded by attrition bias. Another important consideration is that none of the studies on cosmesis followed patients beyond 3 months. In typical scar formation, at the 3-month point, scar width is generally wider than the original incision and erythema is still present beyond the boundaries of the incision. Since scar maturation tends to occur over a period of 1 year, longer follow-up periods in future studies would be needed to determine whether the immediate impact of improved cosmesis using the surgical robot persists.

Our results should be viewed in the context of several potential limitations. First, the cohorts being compared in this meta-analysis were significantly different in many aspects: the patients in the robotic group were younger, had a lower BMI, and proportionally had fewer total thyroidectomies done in comparison to the open group. There are possible differences between American and South Korean patients in terms of body habitus, personal and physician preferences, and so on that may have affected patient selection within the studies we examined in our systematic review and meta-analysis. However, as a whole, the patients in the robotic cohort appear to be lower-risk surgical patients compared with the open cohort.

Methodologically, the current study is at risk of being affected by publication bias, since it is composed primarily of small observational studies demonstrating positive results, and studies with positive results are more likely to be published than those with negative results. Statistical heterogeneity is another concern; 5 of the 11 quantitative comparisons we performed had \( I^2 > 75\% \), which is suggestive of high heterogeneity. This likely reflects the relatively few studies included in our meta-analysis, as well as fundamental differences in the populations being studied within each publication (eg, unilateral vs bilateral thyroid surgery, transaxillary vs combined transaxillary-breast approaches). On the other hand, our largest individual meta-analysis on patient outcomes included only 7 individual studies.
(temporary RLN injury), and such small numbers of studies may be too low powered to reflect true heterogeneity. Thus, caution is advised when evaluating the impact of high heterogeneity on the results of the meta-analysis.

Last, a valid cost analysis remains a key gap in the current literature. We identified 2 studies describing cost but were excluded from analysis because of methodological flaws. Broome et al showed a 217% increase in cost using robotic techniques, but the robotic thyroidectomy cost data were based on published series done at other institutions. Consequently, the cost of robotic surgery may not be reflective of the study institution itself. Cabot et al found significantly higher costs in robotic thyroidectomy compared with other approaches but only evaluated cost for the South Korean data in the study. Because most studies in this meta-analysis were conducted in South Korea, where reimbursement for robotic thyroidectomy is quadruple that for open thyroidectomy, the implications for US-based reimbursement strategies are uncertain. Robotic surgery also generates higher costs from related equipment usage, prolonged operative time, and associated facility and staffing fees. Thus, in the context of escalating health care costs in the United States, this represents a major concern at the national policy level.

Conclusion
Transaxillary and axillo-breast robotic thyroidectomy approaches have complication rates similar to the conventional transcervical approach. However, these robotic approaches also introduce unique complications that should not be ignored. Robotic thyroidectomy has been reported to improve cosmetic outcomes, although long-term studies using validated instruments will be essential to further demonstrate the value of the robotic approach. Valid, well-designed studies on the comparative costs of robotic and open thyroidectomy remain a critical gap in the research literature.

Acknowledgments
The authors thank Christine Landry, MD, Nancy Perrier, MD, and Kyung Tae, MD, for providing supplemental institutional data on robotic and open thyroidectomy patient outcomes and Ron Kuppersmith, MD, for guidance on study design and critical review of our manuscript.

Author Contributions
Gordon H. Sun, substantial contributions to conception and design, acquisition of data, analysis and interpretation of data, drafting the article and revising it critically for important intellectual content, final approval of the version to be published; Lilia Peress, substantial contributions to conception and design, acquisition of data, analysis and interpretation of data, drafting the article and revising it critically for important intellectual content, final approval of the version to be published; Melissa A. Pynnonen, substantial contributions to conception and design, analysis and interpretation of data, drafting the article and revising it critically for important intellectual content, final approval of the version to be published.

Disclosures
Competing interests: The Robert Wood Johnson Foundation and the Department of Veterans Affairs were not directly involved in study design, data acquisition and interpretation, or manuscript preparation or review. Any opinions expressed herein do not necessarily reflect the opinions of the Robert Wood Johnson Foundation or the Department of Veterans Affairs. At the time of manuscript preparation, Gordon H. Sun was a Robert Wood Johnson Foundation Clinical Scholar supported by the US Department of Veterans Affairs. Dr Sun is now an employee of Partnership for Health Analytic Research, LLC in Beverly Hills, California, and a general otolaryngologist at the UCLA Arthur Ashe Student Health & Wellness Center in Los Angeles, California.


Funding source: None.

Supplemental Material
Additional supporting information may be found at http://oto.sagepub.com/content/by/supplemental-data

References


