

Investigating the Sources of Heterogeneity in Systematic Review of Recurrence Rate after Hemithyroidectomy for Low-Risk Well-Differentiated Thyroid Cancer

Axel Sahoalera^{a, b} Marco Ferrari^{b, c, d} Jonathan Athayde^a Adrian Mendez^a
Kevin Fung^a John Yoo^a Anthony C. Nichols^a S. Danielle MacNeil^a

^aDepartment of Otolaryngology-Head and Neck Surgery, Western University, London, ON, Canada; ^bGuided Therapeutics (GTx) Program, Techna Institute, University Health Network, Toronto, ON, Canada; ^cUnit of Otorhinolaryngology – Head and Neck Surgery, Department of Medical and Surgical Specialties, Radiologic Sciences, and Public Health, University of Brescia, Brescia, Italy; ^dSection of Otorhinolaryngology – Head and Neck Surgery, Department of Neurosciences, University of Padua, Padua, Italy

Dear Editor,

The recently published “Systematic Review of Recurrence Rate after Hemithyroidectomy for Low-Risk Well-Differentiated Thyroid Cancer” (WDTC) by Chan et al. [1] aimed to focus on tumors between 1 and 4 cm, which represent the majority of thyroid malignancies. After performing a meta-analysis of the reports which fulfilled their inclusion criteria and using odds ratios (OR), they found higher recurrence rates after hemithyroidectomy (HT) compared with total thyroidectomy (TT), and no significant difference in 10-year survival rates between both approaches.

They acknowledged that the high heterogeneity, I^2 , which is considered significant over 50% [2], in their forest plots hampered the significance of their findings. After a careful reading of the manuscript we endeavored to investigate the sources of heterogeneity and provide an additional analysis of their comprehensive analysis.

The pooled OR in the recurrence rate forest plot which included 25 papers was 1.45 and CI 1.16–1.81 ($p = 0.001$) favoring TT, but the I^2 was 70% ($p < 0.00001$), denoting a heterogeneous group of studies. Subsequently, they identified articles which defined a low-risk cohort of patients

using a recognized risk classification (ATA, AMES, and AJCC) [3–6] and reported recurrence rates in this subgroup (9.2% for HT and 5.3% for TT). We decided to group these 4 papers separately, hypothesizing that the studies using a recognized risk classification should be more comparable if grouped separately. We still found a significant increased recurrence after HT (OR 2.04; CI 1.00–4.14, $p = 0.05$), but in this case the I^2 was 50% ($p = 0.11$), confirming our assumption. Even though the AMES score is still recognized by the British and American Thyroid Associations as a valid classification system, we went even further and performed a subgroup analysis removing the only study which utilized the AMES system [4]. This was done as this score includes tumors <5 cm (AJCC and ATA threshold is 4 cm), which might have increased the clinical heterogeneity grouping of these higher-risk patients with a lower-risk cohort of patients (1–4 cm tumors). Noteworthy, the AMES score is still recognized by the British and American Thyroid Associations as a valid classification. In the final model without the AMES score study, the heterogeneity reduced to 26% ($p = 0.26$), and the OR was 1.38 and CI 0.54–3.54 ($p = 0.5$). In addition, the test for subgroup differences was 61.8%

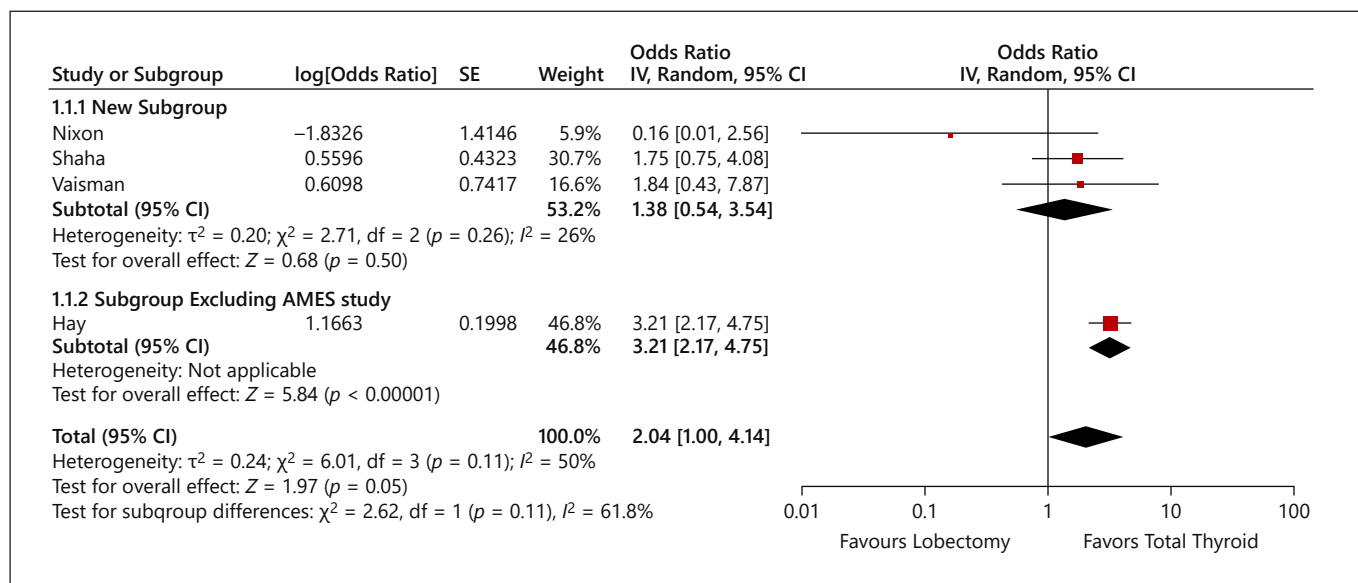


Fig. 1. Forest plot for recurrence in low-risk cohort of patients using a recognized risk classification. OR, odds ratio.

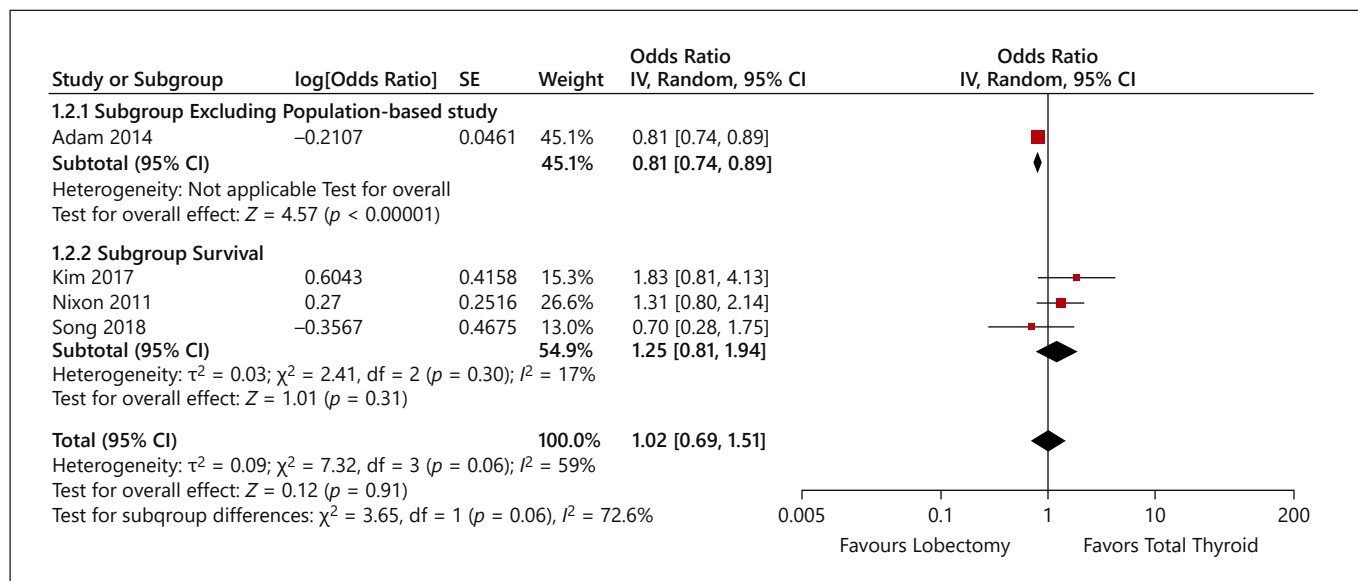


Fig. 2. Forest plot for 10-year overall survival in 1–4 cm well-differentiated thyroid cancers. OR, odds ratio.

indicating that the AMES study may have contributed to the overall heterogeneity (Fig. 1). The definition of recurrence may have also contributed to the heterogeneity. Of the 4 articles included our forest plot, Nixon et al. [6] considered the development of contralateral disease following HT as a second primary tumor, whereas Vaisman et

al. [5] and Hay et al. [4] classified as local recurrence when tumors were present in thyroid remnant. We are more aligned with Nixon et al. [6] regarding this matter and believe that future studies should report recurrences in a more consistent way.

Pooled 10-year overall survival (OS) from 26 articles showed no differences between HT and TT (OR 0.92; CI 0.73–1.18, $p = 0.52$) [1]. Once again, the heterogeneity was high, $I^2 = 85\%$ ($p < 0.00001$), and the authors provided (see online suppl. Table 1; for all online suppl. material, see www.karger.com/doi/10.1159/000511870) organizing studies based on the tumor sizes. We proceeded to strictly extract from this table articles with 1–4 cm tumors, as it represents the size range where less clear recommendations exist to date, obtaining 4 studies [6–9]. The forest plot of these articles showed no differences in OS between approaches (OR 1.02; CI 0.69–1.51, $p = 0.91$), but heterogeneity ($I^2 = 59\%$, $p = 0.06$) was still high. We performed a subgroup analysis removing from the final analysis the only population-based report. These types of studies can sometimes contribute to heterogeneity, one of the reasons being the lack of data granularity. Also, it is highly likely that the population-based study by Adam et al. [8] and the single-institution study by Nixon et al. [6], analyzed same cohort as the first study used US-wide NCBD data with coinciding time periods as the latter study. This would be another reason to perform a sensitivity analysis with and without the population-based study. There was still no difference in 10-year OS (OR 1.25; CI 0.81–1.94, $p = 0.31$), but in this case, $I^2 = 17\%$ ($p = 0.30$). Most importantly, the test for subgroup differences was 72.6% suggesting that the population study was contributing to the overall heterogeneity (Fig. 2).

We capitalized the literature search from Chan et al. [1] and analyzed the heterogeneity in their analyses. In our final models we found that HT approach in low-risk cohort of patients with thyroid cancers was associated with a higher recurrence rate, but in tumors between 1 and 4 cm there appears to be no difference in recurrence rates. For the same size range, we also found that there were no differences in 10-year overall survival between HT and TT. The I^2 scores were nonsignificant, but notably, we were able to identify sources of heterogeneity.

Unquestionably, the main limitation of our analysis is the low number of studies included in our final models. I^2 scores can be biased in small meta-analysis, as they can overestimate the I^2 in homogeneous studies and underestimate it in heterogeneous studies [10]. This mandates that our conclusions should be interpreted cautiously. To illustrate this we calculated the 95% CI as indicated by Higgins et al. [11] for one of our I^2 analysis (first subgroup in Fig. 1) and found wide CIs ($I^2 = 26\%$ CI 0–72.7) indicating that the low number of studies might have af-

ected the reliability of our I^2 scores. Still, we methodologically assessed differences in tumor size and overlapping cohorts in the comparisons identifying them as possible sources of heterogeneity. Other possible sources of heterogeneity that we identified included variable definition of recurrence, whether radioactive iodine (RAI) was administered for HT and varying doses of RAI, but it was not possible to extract this data from the included articles.

HT has been suggested as an alternative to TT as it has lower complication rates including recurrent laryngeal paralysis and hypothyroidism. For nodules 1–4 cm in size, the ATA recommendations are not clear, and HT and TT are possible options for treatment. Therefore, as TT is associated with increased complication rates and morbidity postoperatively, a less invasive alternative would be preferable if it had equivalent treatment outcomes in this size range. Chan et al. [1] endeavored to elucidate this question performing an exhaustive systematic review and meta-analysis, but their conclusions were limited by heterogeneity in the final analyses. Our findings suggest that in patients with 1–4 cm thyroid cancers, classified as low risk by ATA and AJCC scores, recurrence rates may not significantly increase after HT. Moreover, in these size range, non-population-based studies indicate that 10-year overall survival may not be affected by the extent of the approach.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

This study was self-funded.

Author Contributions

All authors provided substantial contributions to the conception and design of the work, and the acquisition, analysis, and interpretation of data. All of them revised it critically for important intellectual content and finally approved the version to be published.

References

- 1 Chan S, Karamali K, Kolodziejczyk A, Oikonomou G, Watkinson J, Paleri V, et al. Systematic review of recurrence rate after hemithyroidectomy for low-risk well-differentiated thyroid cancer. *Eur Thyroid J*. 2020; 9(2):73–84.
- 2 Higgins JP, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ*. 2003;327(7414):557–60.
- 3 Takes RP, Strojan P, Silver CE, Bradley PJ, Haigentz M, Wolf GT, et al. Current trends in initial management of hypopharyngeal cancer: the declining use of open surgery. *Head Neck*. 2012 Feb;34(2):270–81.
- 4 Hay ID, Grant CS, Bergstralh EJ, Thompson GB, Van Heerden JA, Goellner JR, et al. Unilateral total lobectomy: is it sufficient surgical treatment for patients with AMES low-risk papillary thyroid carcinoma? *Surgery*. 1998; 124(6):958–6.
- 5 Vaisman F, Shaha A, Fish S, Michael Tuttle R. Initial therapy with either thyroid lobectomy or total thyroidectomy without radioactive iodine remnant ablation is associated with very low rates of structural disease recurrence in properly selected patients with differentiated thyroid cancer. *Clin Endocrinol*. 2011; 75(1):112–9.
- 6 Nixon IJ, Ganly I, Patel SG, Palmer FL, Whitcher MM, Tuttle RM, et al. Thyroid lobectomy for treatment of well differentiated intrathyroid malignancy. *Surgery*. 2012; 151(4):571–9.
- 7 Song E, Han M, Oh HS, Kim WW, Jeon MJ, Lee YM, et al. Lobectomy is feasible for 1–4 cm papillary thyroid carcinomas: a 10-year propensity score matched-pair analysis on recurrence. *Thyroid*. 2019;29(1):64–70.
- 8 Adam MA, Pura J, Gu L, Dinan MA, Tyler DS, Reed SD, et al. Extent of surgery for papillary thyroid cancer is not associated with survival: an analysis of 61,775 patients. *Ann Surg*. 2014; 260(4):601–7.
- 9 Kim MJ, Lee MC, Lee GH, Choi HS, Cho SW, Kim SJ, et al. Extent of surgery did not affect recurrence during 7-years follow-up in papillary thyroid cancer sized 1–4 cm: preliminary results. *Clin Endocrinol*. 2017;87(1):80–6.
- 10 von Hippel PT. The heterogeneity statistic I2 can be biased in small meta-analyses. *BMC Med Res Methodol*. 2015 Dec 14;15(1):35. : <https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-015-0024-z>. <http://dx.doi.org/10.1186/s12874-015-0024-z>.
- 11 Higgins JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med*. 2002 Jun 15;21(11):1539–58.