Prognostic Value of Lymph Node Yield and Density in Head and Neck Malignancies

Shayan Cheraghlou1, Michael Otremba, MD1,2, Phoebe Kuo Yu, MD1, George O. Agogo, PhD3, Denise Hersey, MLS, MA4, and Benjamin L. Judson, MD1,2

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

Abstract

Objective. Studies have suggested that the lymph node yield and lymph node density from selective or elective neck dissections are predictive of patient outcomes and may be used for patient counseling, treatment planning, or quality measurement. Our objective was to systematically review the literature and conduct a meta-analysis of studies that investigated the prognostic significance of lymph node yield and/or lymph node density after neck dissection for patients with head and neck cancer.

Data Sources. The Ovid/Embase, Ovid/Embase, and NLM PubMed databases were systematically searched on January 23, 2017, for articles published between January 1, 1946, and January 23, 2017.

Review Methods. We reviewed English-language original research that included survival analysis of patients undergoing neck dissection for a head and neck malignancy stratified by lymph node yield and/or lymph node density. Study data were extracted by 2 independent researchers (S.C. and M.O.). We utilized the DerSimonian and Laird random effects model to account for heterogeneity of studies.

Results. Our search yielded 350 nonduplicate articles, with 23 studies included in the final synthesis. Pooled results demonstrated that increased lymph node yield was associated with a significant improvement in survival (hazard ratio, 0.833; 95% CI, 0.790-0.879). Additionally, we found that increased lymph node density was associated with poorer survival (hazard ratio, 1.916; 95% CI, 1.637-2.241).

Conclusions. Increased nodal yield portends improved outcomes and may be a valuable quality indicator for neck dissections, while increased lymph node density is associated with diminished survival and may be used for postsurgical counseling and planning for adjuvant therapy.

Keywords
head and neck cancer, lymph node yield, lymph node ratio, systematic review, prognostic value

Tumors of the head and neck represent a significant disease burden in the United States, with the American Cancer Society predicting >63,000 new cases and >13,000 deaths in 2017.1 While treatment regimens vary by cancer stage and primary site, the majority of cases are treated with either definitive surgery or surgery with adjuvant therapy.2 There is also a growing use of systematic therapy in cases with high-risk pathologic features.3,4 In addition to surgery at the primary site, national guidelines recommend therapeutic or elective neck dissection based on the presence and risk of nodal metastases.5 The operative scope of neck dissections has become more limited in the past several decades, with the role of radical dissections becoming significantly diminished. Furthermore, the use of selective neck dissections and removal of lymph nodes from a subset of levels in the neck rather than from all levels (I-V), particularly in the elective setting, has increased. These more limited procedures have demonstrated outcomes similar to those of more extensive dissections in select patient groups.6-9

Lymph node levels that should be removed in a neck dissection are based on the primary tumor site and are relatively well established; however, there is no consensus on the number of nodes required to constitute a sufficient neck dissection. Notably, there is significant variability in lymph node yield (LNY) based on surgeon technique or...
preference.\textsuperscript{10-12} While it is possible that such variability in LNY reflects inconsequential differences in operative techniques, decreased nodal yield may diminish the likelihood of removing occult nodal disease, thereby leading to poorer survival. Indeed, a number of studies published in the past year suggest this to be the case.\textsuperscript{13-15} LNY is also noteworthy given its potential value as a quality metric for head and neck cancer surgery.

LNY is also used in the calculation of a closely related value, lymph node density (LND), defined as the ratio of the number of positive nodes to the LNY. While LNY carries significance as a potential quality metric, LND may be utilized as a prognostic factor after neck dissection independent of N stage.\textsuperscript{16-18} In fact, some argued that tumor depth combined with LND performs better than standard TNM staging in predicting survival for head and neck malignancies.\textsuperscript{18} A number of studies suggested that in addition to risk stratification, LND may be used as a parameter in the consideration of adjuvant therapies.\textsuperscript{19-21} Others, however, posited that the mathematically calculated nature of LND undermines its prognostic value, leaving it too easily affected by surgical technique.\textsuperscript{22}

The objective of this study was to (1) review and summarize the literature pertaining to the prognostic value of LNY and LND in cancers of the head and neck and (2) determine the hazard associated with differences in LNY and LND through the use of meta-analysis. We reviewed studies reporting outcomes by LNY and LND for all head and neck malignancies or for specific head and neck subsites: nasopharynx, oral cavity, oropharynx, hypopharynx, and larynx. To preserve comparability among studies, we included those that reported overall survival—the most commonly used outcome measure.

**Methods**

**Search and Screening Methodology**

On January 23, 2017, a senior medical librarian performed a comprehensive search of multiple databases: Ovid/Medline (1946–January 2017 [week 3]), Ovid/Embase (1974–January 23, 2017), and NLM PubMed for unindexed materials. We used controlled vocabulary words and synonymous free text words for the topic of interest. In each database, we performed scoping searches and used an iterative process to translate and refine the searches. Key articles, as identified through scoping searches, were used for validating the success of the searches. Our Ovid/Medline and Ovid Embase search strategies are presented in Appendixes A and B, respectively (available in the online version of the article). Our NLM PubMed Search was performed with the following search string:

\[(cancer* OR carcinoma* OR neoplasm*) AND ("head and neck" OR oral OR oropharynx* OR larynx* OR hypopharynx* OR nasopharynx* OR mouth OR gingival OR lip OR tongue OR palat* OR orotrinalaryng* OR lingual* OR tonsil*) AND (yield OR count OR distribution OR examin* OR number OR amount OR density OR ratio) AND (node OR nodes OR nodal) AND (neck dissection) AND ((publisher[sb] NOT pubstatusnihms NOT pubstatuspmcsd NOT pmcbook) OR inprocess[sb] OR pubmednotmedline[sb] OR oldmedline[sb] OR ((pubstatusnihms OR pubstatuspmcsd) AND publisher[sb]))\]

The final search retrieved a total of 449 references, which were pooled in EndNote and then uploaded to Covidence\textsuperscript{23} for screening. Two reviewers (S.C. and M.O.) assessed the eligibility of citations in a standardized manner. Citations were included if they were original investigations and were published in English, if they studied upper aerodigestive tract malignancies, and if they reported overall survival associated with LNY and/or LND in a multivariate analysis. We utilized overall, rather than cancer-specific, survival due to its more ubiquitous availability in cancer registries and other data sources often utilized by researchers. After removal of duplicate records, the abstract of each citation was screened for relevance to the role of LNY and/or LND in head and neck cancers. Irrelevant citations, case reports, and reviews were excluded. Studies of LND were excluded if the reference group was LND = 0, as this would introduce node-negative cases into the analysis of a variable designed to risk-stratify node-positive cases. Full texts were then retrieved for the remaining citations for a final screen by both reviewers. References of the articles meeting the inclusion criteria were reviewed to ensure comprehensiveness.

**Data Extraction and Statistical Analysis**

Data extraction variables were defined a priori. The outcomes extracted from studies of LNY were the hazard ratio (HR) estimate and 95% CI associated with increased nodal yield. The outcomes extracted from studies of LND were the HR estimate and 95% CI associated with increased node density. Where the HR was reported with the higher nodal yield or node density as the reference variable, it was converted into its equivalent for the reverse association. The level of evidence was determined by criteria from the Oxford Centre for Evidence-Based Medicine.\textsuperscript{24}

Because the LNY and LND are related to the extent of disease, we utilized HRs from multivariate Cox regression analyses from each study to control for other disease factors contributing to survival. While this has been done in other meta-analyses,\textsuperscript{25-27} this method has the potential to increase study heterogeneity due to potential differences in adjusted covariates among studies. However, like most of these previous studies, we used a conservative random effects model to account for this heterogeneity. Note that while this model can account for study heterogeneity—in this case, because of differences in the adjusted covariates and the multiple primary sites—our meta-analysis can account only for factors that were represented in the individual multivariate models of the included studies. While a number of critical factors, such as patient age and tumor stage, were routinely cited in these models, many of the studies did not account
for other prognostic factors, such as smoking status, human papillomavirus (HPV) positivity, or the presence of specific tumor markers.

Additionally, given the continuous nature of LNY and LND, different studies utilized different discretization points to delineate between high and low LNY or LND. To address this issue, we excluded outlier discretization points according to a commonly utilized statistical method excluding values (1) less than the first quartile minus 1.5 times the interquartile range or (2) greater than the third quartile plus 1.5 times the interquartile range. Where studies reported separate results for different cutoff values, we included the data for the cutoff closest to the median cutoff among all studies. Studies that reported separate HRs for different cohorts or different LNY/LNR strata had each entry separately included in the analysis, controlling for population size with the weighting method that follows. Finally, we performed a metaregression to identify whether, among the remaining values, differences in discretization was associated with differences reported in hazard values.

Meta-analyses were conducted with the DerSimonian and Laird random effects model. When weighting study effects, this model accounts for study sample size and the estimate of study heterogeneity. HRs were log transformed to obtain effect sizes for inclusion in the model. Standard errors were calculated with the upper and lower bounds of the 95% CI for the HR reported in each study. We derived a \( P \) value for heterogeneity and \( I^2 \) statistic with the Mantel-Haenszel model. Data analysis was performed with STATA 13 (StataCorp LP, College Station, Texas).

**Results**

The PRISMA flow diagram for article selection is presented in Figure 1. A total of 560 articles were identified through 3 computerized database searches (Ovid/Medline, Ovid/Embase, and PubMed). This resulted in 350 articles that went on to abstract screening once duplicates were removed. Of these, full texts were reviewed for 67 articles, and 23 studies were ultimately included in the review.

Tables 1 and 2 report characteristics for studies of LNY and LND, respectively. All studies had a retrospective cohort design. The level of evidence was 2c for 15 of 16 LND studies and 6 of 7 LNY studies. One LNY and 1 LND study had levels of evidence of 2b. Two LNY studies comprised all head and neck cancers, while the remaining 5 included only specific sites. Two LND studies reported all head and neck cancers, while 14 indicated specific subsites. Two LNY studies were conducted at a single institution and 2 across \( \geq 2 \) institutions, and 3 utilized national registry data. Four were conducted within the United States, while 3 were conducted abroad. Thirteen LND studies were conducted at a single institution and 2 across \( \geq 2 \) institutions, and 1 utilized national registry data. Four of these studies were conducted within the United States, while 12 were conducted abroad. The discretization point for LNY varied between 16 and 36, with the majority of studies choosing a cutoff value of 18. One data point was excluded from meta-analysis owing to its discretization point being an outlier (>34.5). The discretization point for LND varied between 1% and 20%, with the majority of studies choosing a cutoff value between 2.5% and 10%. Three data points were excluded from meta-analysis owing to their discretization points being outliers (>13.5%). Metaregression revealed that differences in discretization points were not associated with differences in reported hazard for LNY (\( P = .325 \)) or LND (\( P = .837 \)). Four studies of LNY cited all N stages; 1 study, only cN0 cases; 1 study, only pN0 cases; and 1 study, N0-2c cases. Five studies of LND included all N stages; 2 studies, N0-2c cases; 7 studies, N1-2c cases; and 2 studies, N1-3 cases.

Meta-analysis performed on 7 LNY studies demonstrated a combined HR of 0.833 (95% CI; 0.790-0.879) associated with higher LNY (Figure 2). The \( P \) value for overall effect was <.001. The \( P \) value for heterogeneity was .272, and \( I^2 = 19.2\% \). Meta-analysis performed on 14 LND studies demonstrated a combined HR of 1.916 (95% CI, 1.637-2.241) associated with higher LND (Figure 3). The \( P \) value for overall effect was <.001. The \( P \) value for heterogeneity was <.001, and \( I^2 = 63.4\% \).

**Discussion**

**Lymph Node Yield**

The literature strongly suggests that higher LNYs for therapeutic or elective neck dissections are associated with improved long-term outcomes. There is also some evidence that increased LNY leads to improved survival even when all dissected nodes are negative. However, some discrepancy exists regarding the extent of dissection that constitutes an adequate yield. In one of the earliest studies exploring this association, Ebrahimi et al determined that a nodal yield of at least 18 optimized outcomes for clinically node-negative oral cavity cancers. This figure was derived by stratifying nodal yield into quartiles (1-17, 18-23, 24-31, \( \geq 32 \)) and then determining that while individuals with a nodal yield of 1 to 17 had the poorest outcomes, those with a yield \( \geq 18 \) had similar outcomes. Most studies since have used this cutoff of 18 in studying whether LNY affects survival. A study from our group found that the optimal threshold for LNY for cN0 oral cavity cancers was 16. This was done by grouping cases into 8 strata, 5 nodes each (1-5, 6-10, etc), and determining the lowest LNY at which a significant survival advantage occurred. This method had the advantage of greater accuracy than what could be achieved by using quartiles, since each group contained a smaller number of values. This work was also the first to study the optimal LNY threshold for cN+ cancers, and found that this value was higher than that of cN0 cancers, at 26.

Regardless of the exact threshold that is ultimately reached for LNY during neck dissection, it is becoming increasingly clear that this measure will be an important quality indicator for operations involving therapeutic or elective neck dissections. LNY may provide an objective value that institutions and individual surgeons can use to...
assess the impact of quality improvement projects, operative techniques, and surgical instruments on their quality of care. It has been well established throughout medicine that objective quality metrics lead to improved provider performance and patient outcomes. Indeed, as a modifiable treatment factor that significantly affects patient survival, LNY as a quality metric may very well improve long-term outcomes for individuals with head and neck malignancies. Additionally, the more widespread achievement of adequate nodal yields may aid in more accurate staging of patients with head and neck cancer. In the new staging criteria of the American Joint Committee on Cancer (AJCC; eighth edition), cancers are staged differently according to the number of recovered positive lymph nodes. For example, pN1 and pN2 HPV-positive oropharynx cancer is differentiated on the basis of whether there are >4 lymph node metastases. Without adequate neck dissections, if positive nodes are missed, patients with >4 positive nodes may be inaccurately staged as pN1.

However, as with any objective measurement of a nuanced surgical or clinical process, it is prone to limitations. Most apparent of these is the influence of individual patient factors as well as disease factors on nodal yield. Studies of LNY from other tumor sites demonstrated that factors such as patient age and weight, as well as tumor size, stage, anatomic location, and grade, influence the LNY after lymph node dissection. Previous radiation has also been shown to affect reported LNY. Additionally, a number of factors influence the number of nodes identified by pathology: between-pathologist variation, between-laboratory variation, and pathologist training level. Thus, while LNY may be a strong summary statistic for surgical quality, it is important that patient and disease factors be adjusted for and pathology procedures be standardized before this is done.

Figure 1. PRISMA flow diagram.

Cheraghliou et al 1019
Lymph Node Density

It is clear from the literature that high LND in the neck is associated with poor prognosis for patients with head and neck cancers. This has been shown in studies of oral cavity cancers, hypopharyngeal cancers, laryngeal cancers, oropharyngeal cancers, and tongue neoplasms, as well as studies of all head and neck subsites. In fact, only 1 study in this review concluded that LND was not a significant prognostic indicator, and it did so for laryngeal cancer specifically.

More so than the relatively smaller set of studies of LNY, those of LND varied widely on the cutoff values used to determine high versus low or intermediate LND. Despite omitted outlier values and the finding that the study cutoff value did not significantly influence the reported hazard, these values ranged from 1% to 20%. The most robust study of the appropriate cutoff point to use for LND was done in oral cavity cancers by an international consortium, which settled on a cutoff value of 7%.

The implications of the prognostic significance of LND are substantially different from those for LNY. Rather than measuring a modifiable treatment factor, LND is more representative of the extent of disease. Thus, the value of LND...
lies in postsurgical patient counseling as well as planning of adjuvant therapy. Many have called for the addition of LND as a complement to standard TNM staging, with some going so far as to argue its superiority over this more standard staging system. Nevertheless, despite its prognostic significance in malignancies of multiple primary sites, including a number outside the head and neck, the eighth edition of the AJCC staging manual does not cite LND in its criteria for N categorization. While distinctions are made between cancers with spread to a single ipsilateral node and multiple ipsilateral nodes and, in the case of HPV-positive oropharynx cancers, between cancers with $\leq 4$ and $>4$ lymph node metastases, these definitions do not account for the number of nodes that were examined for staging. Based on our findings from the literature, future consideration of LND in these definitions may increase the prognostic accuracy of tumor staging. A number of the studies reviewed also point to the value of LND in risk-stratifying patients to inform decisions surrounding adjuvant therapy. However, the literature lacks evidence of a demonstrable difference in benefit of adjuvant treatment based on LND.

The value of LND also suffers from some of the same limitations as LNY. Given that LNY is the denominator in the measure of LND, factors that affect LNY, such as pathologist/laboratory variability, will also affect the measured LND. Use of LND as a prognostic indicator on an individual patient level is also problematic given its fairly stochastic nature. Even with a fully accurate reading of the neck dissection sample, the density of nodes in the sample may not be representative of the density of nodes in the neck. For example, in a hypothetical neck with 50 clinically negative nodes of which 5 harbor occult disease, the true LND of the neck is 10%. However, a surgeon may remove anywhere from 0 to 5 of the positive nodes, which could result in a measured LND anywhere from 0% to 25%, if we assume an LNY of 20—the approximate minimum acceptable level discussed earlier. On the basis of these many issues, some have suggested using the number of positive nodes as a prognostic indicator, instead of LND. Indeed, a recent study demonstrated that the number of positive nodes was a stronger prognostic indicator than both LND and AJCC (seventh edition) N staging for squamous cell cancers of the head and neck.

**Conclusions**

LNY and LND are both associated with differences in survival after neck dissection for malignancies of the head and neck, with increased LNY and decreased LND being associated with improved survival. While there are some limitations to the measure, LND may be used in conjunction with AJCC TNM staging to risk-stratify patients after surgery. LNY is a modifiable treatment factor that affects patient outcomes and may be increasingly utilized as a quality metric for neck dissections. However, before more widespread implementation of this measure, important modifying factors, such as patient disease and pathologist variability, must be accurately adjusted for.

**Author Contributions**

Shayan Cheraghlou, conception and design, analysis and interpretation, drafting manuscript, primary reviewer; Michael Otremba, conception and design, revision of manuscript, primary reviewer;
Phoebe Kuo Yu, conception and design, revision of manuscript; George O. Agogo, conception and design, revision of manuscript; Denise Hersey, conception and design, structuring of search methods, revision of manuscript, drafting manuscript; Benjamin L. Judson, conception and design, interpretation of data, revision of manuscript.

Disclosures

Competing interests: None.
Sponsorships: None.
Funding source: Yale School of Medicine Lowe Endowment, Yale School of Medicine Taylor Opportunity Student Research Fellowship.

Supplemental Material

Additional supporting information is available in the online version of the article.

References


