High Variability in Lymph Node Counts Among an International Cohort of Pathologists: Questioning the Scientific Validity of Node Counts

John P. Sherbeck, MD; Lili Zhao, PhD; and Richard W. Lieberman, MD

Abstract
Background: The enumeration of lymph nodes (LNs) from surgical specimens plays a critical role in the staging of patients with cancer. LN count (LNC) can affect prognosis, staging, adequacy of resection, and/or eligibility for clinical trials. However, there is no standard method for counting LNs. Most studies in the literature site the pathology report as the source of LN data, without discussion of the counting criteria. Patients and Methods: Four microscopic slides from separate pelvic LN dissections were digitally scanned and uploaded with their gross descriptions to an online library and an online survey. Respondents were asked how many LNs they would count per slide as part of a staging procedure. The survey was distributed to an international cohort of pathologists. Results: A total of 122 surveys were returned: 79 from practicing pathologists and 43 from residents/fellows. There was no statistical difference between the groups. All slides showed significant individual variability. The LNC range for each slide was as follows: slide 1, 1–3; slide 2, 0–13; slide 3, 1–8; slide 4, 1–11. The intraclass correlation (ICC) for all responders was 0.26 (95% CI, 0.05<ICC<0.74), which demonstrates a very low agreement among individuals. Although there is a small amount of literature assessing causes of variability in counts, nearly all focus on different techniques at the grossing bench. Our study is the largest prospective assessment of LNC by an international cohort of pathologists. We have demonstrated tremendous variation in the number of LNs pathologists report for a given slide, thus significantly altering how many total LNs are counted. This calls into question the clinical utility of node counts, as well as their use as a quality indicator. Conclusions: LNC is subject to tremendous interpathologist variation, which has a significant clinical and research impact. Consensus in pathologic handling and microscopic enumeration of LNs is essential.

The pathologic evaluation of lymph nodes (LNs) for metastatic disease plays an important role in the surgical staging of nearly all solid-organ malignancies. Nodal tissue is evaluated by pathology either as part of an en bloc resection or as separate, anatomically distinct specimens. In either case, the specimen is grossly examined and LNs are dissected and submitted for histologic examination via a variety of nonstandardized methods. The surgical pathology report typically includes both the total LN count (LNC), as well as the number of positive LNs, and has developed into a critical aspect of planning cancer treatment protocols. According to the 8th edition of the AJCC Cancer Staging Manual, the pathologic node staging (pN) of 29 of the approximately 70 solid-organ staging systems use LNCs in some form and, of these, 13 are staged solely on the number of positive LNs. LNC has been proposed as a metric to assess adequacy of surgical resection and/or pathologic examination, and thresholds of a minimum LNC have been required for inclusion in a number of clinical trials.
Interest regarding LNC and clinical outcomes is abundant. The literature is replete with large studies investigating the role of LNC and outcomes, many of which demonstrate improved outcomes with higher LNCs. However, these studies provide very little information regarding methods of nodal quantitation. Nearly all of the published literature simply cites the “pathology report” as the source of LNC data. In the studies that do give additional information on LNC methods, it is typically limited to a brief description of the tissue-processing technique and a nominal statement that tissue was evaluated by standard microscopy.

The lack of formally described LNC methods in these studies may come as a surprise to most non-pathologists. To the uninitiated, LNs in a specimen could be easily compared with jelly beans in a jar: dump them out and count. However, the actual process involves many factors, each with their own challenges. Multiple studies have attempted to assess LNC variability and have demonstrated a variety of factors, including the patient’s anatomy, tumor biology, experience of the gross examiner, and use of fat-clearing solutions. Although many of these studies also identify the pathologist as a significant source of variation, this has not been uniformly true. By far, the least investigated aspect of LNC is the actual enumeration process of the pathologist during microscopic review; however, this is the time at which the LNC is determined. To date, a single study has evaluated this aspect and demonstrated significant interpathologist variability. However, the study was somewhat limited in the number of pathologists (N=10), and the heterogeneity of their practice setting (2 departments). In order to further assess whether LNC represents a scientifically meaningful number, we conducted the largest and most diverse study of LNC by pathologists.

Patients and Methods
Selection of Slides
Four routine pelvic lymphadenectomy slides from gynecologic surgical procedures at the University of Michigan Department of Pathology were selected for the study in order to represent a sampling of the variability seen in routine lymphadenectomies. Each slide was presented with the associated “gross description” as communicated during specimen processing and extracted from the final pathology report: (1) a single gross LN submitted intact, (2) fibroadipose tissue without definitive gross LNs, (3) single gross LN sectioned into multiple pieces and submitted entirely into a single cassette, and (4) single gross LN bisected and submitted into a single cassette. All slides were negative for carcinoma. Slides were digitally scanned at ×40 magnification, scrubbed of patient identifiers, and uploaded to an in-house digital slide library (Figures 1–4).

Pilot Survey
A pilot survey was created to assess feasibility and refine the process. The pilot was limited to the 4 slides and their gross descriptions, and was distributed via email to pathology trainees (residents and fellows; N=48) at the University of Michigan. The survey simply asked, “For each slide, how many LNs would you count as part of an oncologic staging procedure?”

Main Survey
The second survey was identical to the pilot survey, but included a demographics question to separate trainees from nontrainees, and 3 subjective, multiple-choice questions: (1) “Which features do you require for your minimum definition of a distinct (ie, countable) LN,” (2) “How do you use the gross description in counting LNs,” and (3) “Does the specimen type and/or site impact

Figure 1. Gross description: received in formalin, in a small container, is a 4 x 2 x 1-cm portion of fibroadipose tissue (hematoxylin-eosin, original magnification x1.4). Dissection reveals up to 3 candidate lymph nodes, 1.5 to 2.3 cm in greatest dimension. Slide index: one candidate lymph node submitted whole.
the way you count LN(s)?” Respondents were able to choose more than one answer and to submit free-text comments for each question. The second survey was distributed using a free online survey tool (www.surveymonkey.com), with embedded links to the online slide library. The survey was distributed via email to selected faculty and alumni pathologists of the University of Michigan and members of the International Society of Gynecological Pathologists (N=345) and the British Association of Gynaecological Pathologists (N=180), and ad hoc to pathology colleagues at various institutions. Recipients were encouraged to pass the survey around to any pathologist colleagues who might be interested; therefore, the response rate could not be assessed. All surveys were completed voluntarily and all survey responses were anonymous.

**Statistical Analysis**

Median, mode, and interquartile range for LNC assignment of each slide were calculated (Table 1). Interpathologist agreement was assessed using intraclass correlation (ICC), which is defined as the ratio of the between-slide variance to the total variance (between-slide variance + within-slide variance). ICC assesses similarity of the quantitative measurements (ie, LNCs) among pathologists in multiple, discrete data sets; an ICC close to 1 indicates high similarity. The ICCs and the corresponding 95% CIs were estimated using the linear mixed-effect model in JAGS.

**Results**

**Lymph Node Counts**

In total, 122 surveys were returned, with 120 giving an LNC for all 4 slides. A total of 43 respondents, including 24 from the pilot study, identified as trainees (pathology residents or fellows), whereas 79 identified as practicing pathologists. Response rates could not be assessed given the open nature of the

<table>
<thead>
<tr>
<th>Table 1. Mean, Mode, and Range of Lymph Node Counts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>Mean</td>
</tr>
<tr>
<td>Mode</td>
</tr>
<tr>
<td>Range</td>
</tr>
</tbody>
</table>
survey. All 4 slides demonstrated significant variability in the LNC, as shown in Figure 5. The ICC for all responders was 0.26 (95% CI, 0.05<ICC<0.74), which demonstrates a very low agreement between individuals. No statistical difference in variation was seen between the trainee and nontrainee groups by ICC (0.21 vs 0.27).

**Subjective Multiple-Choice Questions**

A total of 98 surveys were returned with answers to the multiple choice questions (Table 2). Each question had a majority answer pattern. Regarding node definition, 79% indicated that their minimum criterion for a countable LN is a lymphoid aggregate with a capsule. However, 14% required a discretely palpable/grossly identifiable LN, whereas one included free text indicating a minimal size criterion (1.0 mm). Regarding use of the gross description, 73% took into account the gross assessment of LNC, whereas 93% use the gross to help determine whether sectioning of an LN occurred during processing. However, 6% responded that they do not use the gross at all in arriving at their LNC. Notably, 3 responded with free text indicating that the number of LNs reported in the gross is favored over the microscopic count in instances of discrepancy. Interestingly, although 80% responded that the specimen type/site did not influence their counting strategies, 16% responded that a minimum LNC requirement might lead them to count more liberally. Finally, multiple nontrainee respondents included free text acknowledging the challenge of consistent counting, with one acknowledging that when working with residents, they default to the resident’s count.

**Discussion**

Much attention has been paid to the variables in LNC, including the role played by pathology, with a number of studies showing the pathologist to be among the greatest sources of variance. These studies tend to focus on LN specimen processing techniques, and postulate that LNC variance is related more to experience, skill, or diligence. Importantly, these studies are retrospective in design and have variable input by pathologists. This latter point is incredibly important: clinicians are generally unfamiliar with the actual method by which LNC is derived, and thus conclusions drawn solely from their retrospective assessment of the pathology report represent a major limitation of these publications. Excluding the significance of positive LNs in cancer staging, our study’s data shed doubt on the scientific utility of enumerating LNC in gen-

---

**Table 2. Results of Subjective Multiple Choice Survey Questions**

<table>
<thead>
<tr>
<th>LN definition</th>
<th>Respondents (N=98)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lymphoid aggregate</td>
<td>17</td>
</tr>
<tr>
<td>Lymphoid aggregate with capsule</td>
<td>77</td>
</tr>
<tr>
<td>Lymphoid aggregate with follicular architecture</td>
<td>19</td>
</tr>
<tr>
<td>Grossly discrete and/or palpable</td>
<td>21</td>
</tr>
<tr>
<td>Other comments</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Use of gross description</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>I don’t use the gross to count</td>
<td>6</td>
</tr>
<tr>
<td>I use the gross to enumerate how many LNs were grossly identified within a cassette</td>
<td>72</td>
</tr>
<tr>
<td>I use the gross to determine if an LN was submitted whole or sectioned</td>
<td>91</td>
</tr>
<tr>
<td>Other comments</td>
<td>10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Impact of specimen type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Specimen types/site does not impact how I count</td>
<td>78</td>
</tr>
<tr>
<td>I may count more liberally to meet an LNC requirement</td>
<td>16</td>
</tr>
<tr>
<td>My institutional protocols necessitate different counting criteria for different specimens</td>
<td>4</td>
</tr>
<tr>
<td>Other comments</td>
<td>7</td>
</tr>
</tbody>
</table>

Abbreviations: LN, lymph node; LNC, lymph node count.
*More than one answer was permitted for each question.
eral, or as a metric for adequacy of excision or sampling. Importantly, our results prove that the wide variation in LNC begins with the pathologist at the microscope.

Challenges in Pathologic Counting of LNs

Parkash et al\textsuperscript{14} and Khalifa and Smith\textsuperscript{20} summarized the spectrum of challenges, from processing of samples to enumerating the LNC from glass slides. As they point out, the grossing bench is the first source of variation. There is wide variation between laboratories in techniques used, including use of fat-clearing solutions, grossing specimens both fresh and/or after formalin fixation, and wholesale submission of the entire specimen. However, in general terms, the adipose tissue is palpated, then sharply dissected to find LNs. In many cases, this involves complete sectioning of the tissue at regular intervals, introducing a potential source of error: unrecognized sectioning of a single node. Once the dissection is complete, LNs are submitted for processing in cassettes. Multiple small LNs may be submitted within a single cassette, meaning a given slide will have multiple LNs on it. However, it must be emphasized that even the most experienced prosector will mistake various nonnodal tissue for small LNs, including nodular scar tissue, thrombosed vessels, ganglia, and granulomatous inflammation. Because assessment of as many LNs as possible is the goal, inclusion of these “possible” or “candidate” LNs is preferred. As a result, it is paramount that there is gross–microscopic correlation of these specimens.

Once the gross–microscopic correlation of the submitted nodes is complete, the nodes themselves must be re-enumerated for the final report. Although this may seem like a routine, binary process, it is not. The anatomic and functional definition of an LN is well characterized, both in text and in practice. Yet, there is no definition for histologically distinct nodes within a single tissue section or glass slide. This distinction is ultimately at the discretion of the signing pathologist and, as shown in our study, is a significant source of variation. Anecdotally, a given pathologist’s LNC practice is cultural: they count the way their mentors and/or colleagues count. This is evident in studies that demonstrate significant differences in LNC between pathology departments, despite identical surgeons and case mixes.\textsuperscript{8,21} Because LNs can be irregularly shaped, infiltrated by adipose tissue, and arranged in chains, a single tissue piece may include multiple fragments of distinct nodal tissue. Does this represent LNs in a chain? Or is it the result of a plane of sectioning that, if continued, would connect those fragments into one LN? Answering these questions is a technical, or at least practical, challenge.

Further complicating the LNC is the practice of submitting random fragments of vascularized adipose tissue to histology with the hope of increasing the LNC. Examination of these fragments often reveals scattered microscopic LNs. This practice raises 2 significant questions: what minimum histologic criteria should be used to define a clinically significant small LN, and what effect does a required minimal LNC have on pathologists’ practices? Regarding the first question, Markl et al\textsuperscript{22} challenge this practice in the colon cancer literature, stating “…reporting very small negative LNs (≤1 mm) is more a whitewash of the report than a serious report.” A similar study by Euscher et al\textsuperscript{23} demonstrated a similar lack of clinically significant findings in gynecologic procedures. Multiple studies have shown minimum LNC requirements (often required by AJCC staging criteria or cancer treatment trials) to induce the Hawthorne effect: the pathologist’s knowledge of potential scrutiny of their reported LNC will result in higher counts.\textsuperscript{10,21} This finding is supported by our study, in which 16 pathologists state that they have different, more liberal counting criteria in the setting of a required minimum LNC.

Research Implications of Count Variation

Research into LNC and patient outcomes have linked together a significant impact of higher LNC with increased survival.\textsuperscript{4–7,24–27} Most of these studies have been in gynecologic and colorectal cancer, and have demonstrated variable minimal LNCs needed to be placed into the improved outcomes bin. However, essentially none of these studies address LNC in their reported methods. Further, the source of the pathologic material (in-house vs referral) is not uniformly addressed. Although central review of pathologic material is the standard for patients who are transferring care, this process complicates reporting of LNCs from multicenter trials. For the pathologist performing a secondary review (eg, second opinion, referral to tertiary care center for treatment), there are 2 options: report out...
the original LNC or attempt to recount. Neither of these situations are ideal. Failure to perform and report an independent LNC defeats the purpose of secondary review as it relates to LNC. However, variations in LN grossing techniques and gross description nomenclature can make this particularly challenging. Further, the reviewing pathologist may not have access to all of the original material (slides or reports) needed to arrive at a count, because some institutions may not routinely include nondiagnostic slides (ie, those without cancer) when fulfilling requests for outside review.

Thus, although material and data from multicenter trials and inclusion of referral cases are typically seen as favorable in outcomes-based, clinicopathologic research, LNC data that are derived from these may be subject to multiple confounding factors. Realizing this challenge, investigators have expressed the need for “caution” in the interpretation of LNC data.27 Although there is no doubt that numerous retrospective studies have shown improved outcomes based on a certain LNC, the previously mentioned reasons make it difficult to broadly apply their data, and indeed their LNC cutoffs, to individual patients.

Clinical Implications of Count Variation
Despite published reports of the impact of LNC on outcomes, to date there has been limited influence on LN specimen processing and reporting. The College of American Pathologists recommends a minimum LNC (N=12) in colorectal cancer, which has been adopted by the AJCC. For endometrial cancers, although the AJCC does not require a minimum, the widely used FIGO staging criteria requires 6 nodes to be assessed; otherwise the stage is designated as pNX.1 This has significant impact beyond an individual patient’s designation of incomplete staging: trial inclusion. Multiple Gynecologic Oncology Group trials require complete FIGO staging, and thus a minimum LNC (ClinicalTrials.gov identifiers: NCT00002706 and NCT00340808). Complete staging, in and of itself, is certainly a reasonable criterion for maintaining good study control. However, defining complete staging, and thus eligibility, based solely on a highly variable LNC has tremendous repercussions. For the individual patient, participation in a trial may have immense altruistic meaning beyond their own outcome.28 For the institution, a protocol violation for a patient deemed “ineligible” due to inadequate LNC may disrupt the integrity of secondary review or risk the institution’s continued participation in trials. In our data set, the LNC range for a single slide varied from 0 to 12. The effects these 2 extremes have on a final LNC speak for themselves.

It is important to note that, although our study was limited to benign LNs, the same challenges exist in counting and reporting the positive lymph nodes (+LNC), both in determination of the numerator (+LNC) as well as the denominator (LNC). This represents a hurdle in deriving accurate, reproducible nodal staging for a significant number of cancers.

Next Steps
As investigators continue to study the role LNC plays in outcomes, it is essential to pause and recognize the immense variation in LNC. Although it would be beneficial for pathology laboratories to adopt uniform procedures in the gross handling of these specimens, variations in resources make this incredibly challenging. Further, our study demonstrates that even with uniformity of the slides, there is significant variation in pathologists’ LNC practice. This could be mitigated by instituting uniform guidelines and definitions of what makes an LN uniquely countable. However, no such definition exists. Thus, it is important for the medical field as a whole to revisit how LN staging specimens are analyzed, interpreted, and reported.

Our study does not address the increasing importance of sentinel LN sampling as part of a multitude of cancer staging protocols across different systems (ie, vulva, melanoma, breast, and now endometrium). Subsequent therapeutic intervention in these cases invariably relies on the pathology report. As our study infers, LN counting and reporting is variable. Accordingly, a unified effort to standardize sentinel LN sample processing and reporting of histopathologic findings, positive or negative, should begin soon in order to assure that the science of the data gathered can be evaluated without hesitation.

Conclusions
This study demonstrates significant variation in the LNC among an international cohort of pathologists.
As LNC continues to drive clinical decision-making, this study portends a tremendous impact on both bedside care and translational research. These findings highlight the challenges faced by pathologists in arriving at a reliable, uniform count. Although using consistent grossing methods and counting criteria may be a solution, the subjective interpretation required to report an LNC is a high hurdle to overcome. This requires the entire field to step back and look at our metrics for staging, specifically downplaying use of LNC as a metric for adequacy, and come to consensus on a more reliable and scientific way of assessing and reporting LNCs.

Acknowledgments

We would like to thank the International Society of Gynecological Pathologists and the British Association of Gynaecological Pathologists for the distribution of the survey to their members. We are indebted to the gynecologic oncology and radiation oncology faculty at the University of Michigan for their thoughtful input into the manuscript.

References